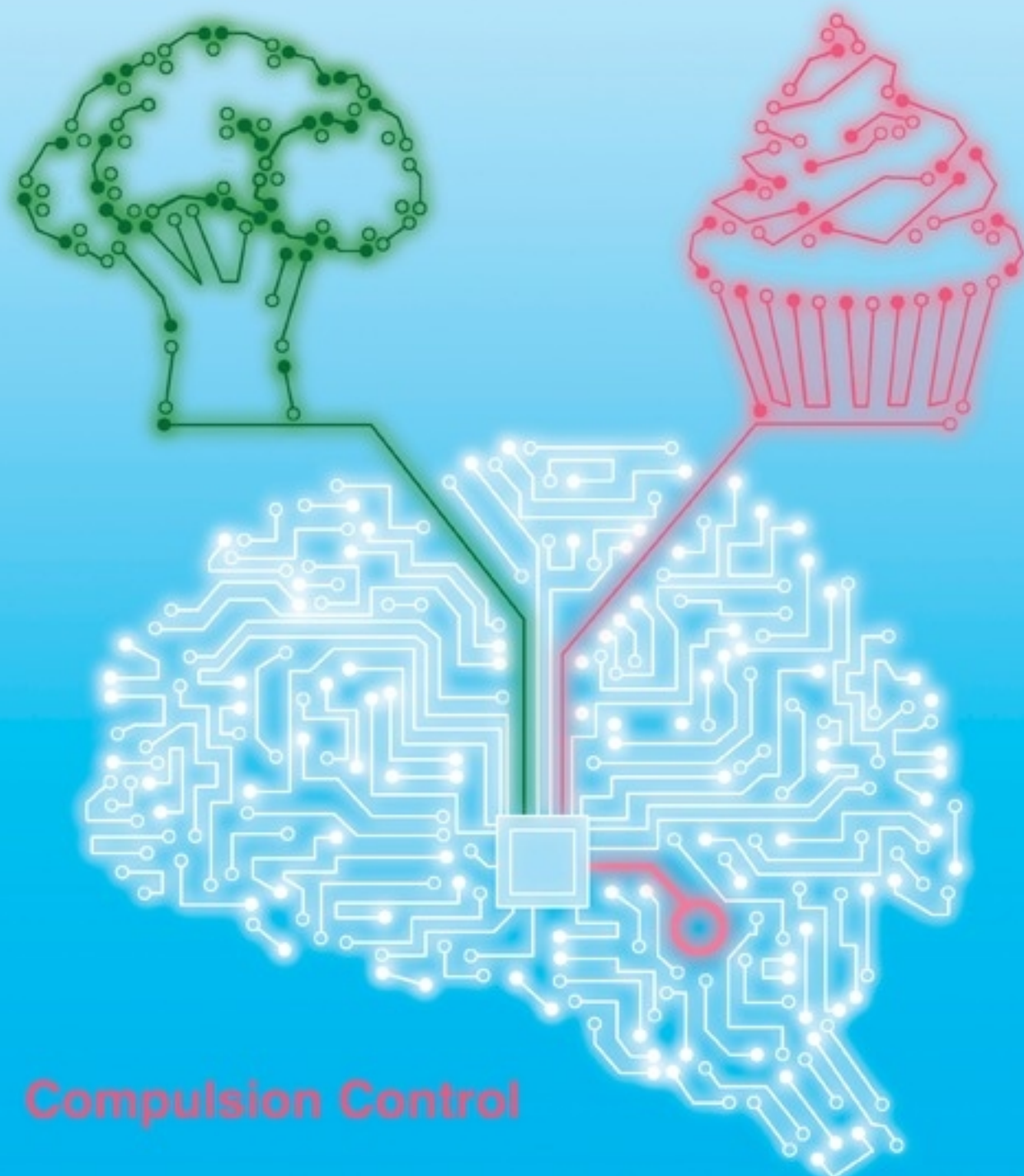


Cell

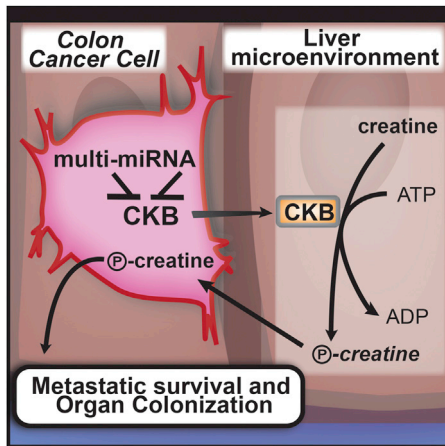
Volume 160
Number 3

January 29, 2015

www.cell.com



Compulsion Control



Tumor Sends Kinase on Food Run

PAGE 393

The tumor microenvironment is a rich source of metabolic substrates that can be utilized by cancer cells. Loo et al. find that colon cancer cells secrete a kinase that acts extracellularly to generate one such metabolite, phosphocreatinine, that directly fuels tumor growth and metastasis to the liver. Inhibiting this kinase with small molecules or adenoviral vectors suppresses metastatic colonization in mice and may prove beneficial in humans.

Transcription: The Director's Cut

PAGE 367

Puc et al. report that Topoisomerase 1 nicking function is required for robust enhancer activation. DNA nicking is necessary to relieve torsional stress and promote eRNA synthesis, revealing an unexpected connection between transcription and the DNA damage repair response.

Signals Entrained Above (and with) the Noise

PAGE 381

Noisiness in molecular systems has been considered detrimental for signal transduction. Now, however, Kellogg et al. show that noise can work in concert with signal oscillations to control the transcriptional activity of NF- κ B. Intrinsic biochemical noise in individual cells promotes oscillation of NF- κ B entrained to fluctuating cytokine signals, and cell-to-cell variability in NF- κ B levels creates population robustness. Together, the two types of transcriptional noise enable signal entrainment over a wider range of dynamic inputs.

Argonaute's Partner in Silencing

PAGE 407

Argonaute proteins typically cleave mRNA targets during RNAi, but their slicing activity is not required for silencing in *C. elegans*. Tsai et al. identify a ribonuclease that is recruited by Argonaute and promotes cleavage and uridylation of RNAi targets, which are then used as templates for siRNA amplification. The findings suggest how Argonautes promote both mRNA silencing and amplification of the silencing signal.

Latency in Quiet Places

PAGE 420

It has been recently proposed that latent HIV often integrates in or near cell-cycle or cancer-related genes, leading to clonal expansion and persistence of CD4⁺ T cells. Now Cohn et al. show that clonally expanded T cells chiefly contain defective proviruses and thus do not form a biologically active latent reservoir. Instead, the reservoir resides in quiescent cells containing single integrations within silent genes or intergenic regions.

Bespoke Antibodies for HIV

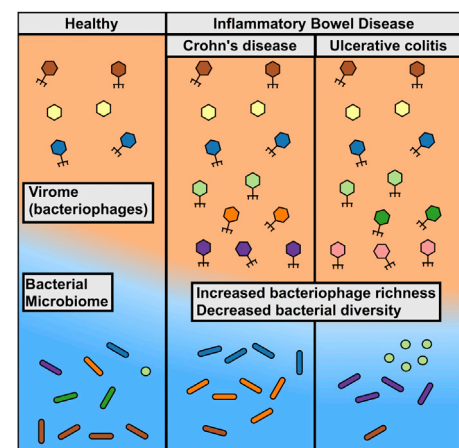
PAGE 433

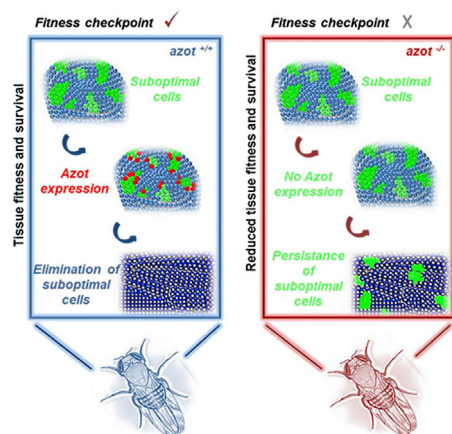
While antibodies generally neutralize viruses by bivalent binding to neighboring virion spikes, HIV-1 spike architecture prohibits intra-spike crosslinking by naturally occurring antibodies. Galimidi et al. develop a class of engineered antibody-based molecules, rationally designed for high avidity intra-spike binding, that overcomes this barrier and presents over 100-fold increase in HIV-1 neutralization potency.

The Inflammatory Virome

PAGE 447

The interplay between bacterial diversity and composition and enteric disease is increasingly understood, yet little is known about the effects of the virome. Norman et al. find that the enteric virome is perturbed in inflammatory bowel disease patients and uncover specific features of the viromes in Crohn's disease and ulcerative colitis that are not explained by changes in bacterial diversity and richness.





Healthy Competition between Cells

PAGE 461

Damaged cells can accumulate during development and aging. Merino et al. now identify a new *Drosophila* gene, *azot*, as the key factor ensuring the elimination of these less fit cells. These findings enable the authors to demonstrate that direct cell-to-cell fitness comparison and selection is essential for maintaining organismal health and extending lifespan.

Mice Not Missing MYC (at all!)

PAGE 477

Hofmann et al. report that reduced expression of MYC increases lifespan in mice and makes their aging tissues healthier too. These benefits occur without apparent developmental trade-offs or significant changes in stress management pathways. Rather, they appear to be mediated by a unique combination of changes in core nutrient and energy sensing pathways.

Kinase-cum-Tumor Suppressor

PAGE 489

Protein kinase C (PKC) inhibitors have been unsuccessful in clinical trials. Bioinformatic, genetic, and biochemical analyses of human cancer-associated PKC mutations by Antal et al. reveals that the majority are loss-of-function, uncovering an unexpected role for PKC as a tumor suppressor. Since reduction of PKC function enhances tumor growth, therapeutic strategies to restore rather than inhibit PKC activity should be pursued.

Light Touch and Go

PAGE 503

Different kinds of sensory touch input (stretch, vibration, heat) are processed by different sets of neurons. Bourane et al. identify a class of spinal cord interneurons essential for transmitting light touch input from receptors in the skin. These neurons additionally integrate sensory touch information with motor control input from the brain to generate the corrective motor movements needed to balance when walking on uneven surfaces.

Seek, Consume, or Binge

PAGE 516 and PAGE 528

Animals need to seek out and then consume food. Jennings et al. define the subset of neurons in the lateral hypothalamus responsible for encoding the two behaviors. Combining optogenetics and calcium imaging to measure neuronal activity at single-cell resolution in freely behaving mice, they find that seeking and consuming behaviors are encoded by distinct neurons, suggesting that the two behavioral aspects can be dissociated. In a related paper, Nieh et al. describe the neural circuit loop that selectively controls compulsive sugar consumption but does not affect feeding necessary for survival, providing a potential target for compulsive overeating therapeutic interventions.

Enhancers on Evolution's Fast Track

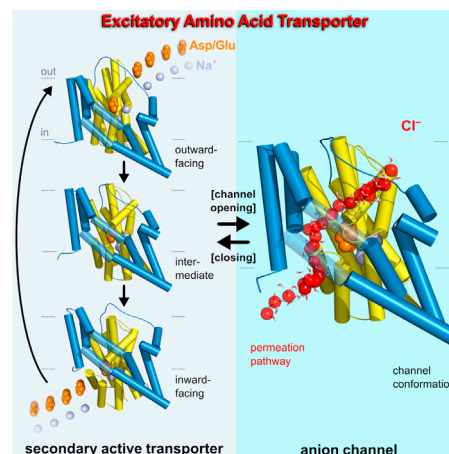
PAGE 554

A comparative functional genomic analysis in liver samples from 20 mammalian species enable Villar et al. to identify that, in contrast to promoters, enhancers evolve rapidly and that recently evolved enhancers dominate the mammalian regulatory landscape. Interestingly, most of these enhancers arise from genomic regions present in the mammalian ancestors, rather than lineage-specific expansions of repeat elements, and they are associated with genes under positive selection.

Channeling a Transporter

PAGE 542

Excitatory amino acid transporters operate both as transporters and as anion-selective ion channels at synapses. Machtens et al. employ a combination of simulations and experiments with prokaryotic and mammalian glutamate transporter homologs to define the anion conduction pathway and elucidate how a class of active transporters can function as selective, gated anion channels.



Building a Superhero

Innovation in the natural world inspired many budding and established scientists, providing insights into species' adaptation and the processes that drive it. More recently, and as our ability to manipulate materials expands, evolutionary inventions have also provided creative solutions for human challenges. The development of Velcro, designed to the likeness of thistle's barbed seeds (de Mestral, 1958), is an example of an invention where the idiosyncratic problem of seed dispersal translated into a universal solution for our need to keep things together.

Nature-inspired designs, or biomimetics, can be also applied to less obvious problems. One recent example is the study from the laboratories of Tae-il Kim and Mansoo Choi describing an ultra-sensitive mechanosensor that can pick up faint physiological outputs such as blood pressure, heartbeat, or voice patterns (Kang et al., 2014). The sensor was inspired by the lyriform organ, a specialized exoskeletal structure found on the limbs of spiders, consisting of parallel slits of different lengths connected to the nervous system. This arrangement, reminiscent of the strings in a lyre and used for communication, allows spiders to pick up exquisitely fine vibrations from their surroundings. Kim and Choi mimicked the lyriform organ by layering nanometers-thick platinum on top of a viscoelastic polymer and generating cracks in the platinum and polymer layers. The cracks are the sensing organ—electrical conductance across them depends on the size of the gap, allowing measurement of the fine vibrations and pressure that distorts the layers. This wearable sensor is indeed able to pick up fine distinctions in speech or blood flow, and one can imagine a wide range of future applications where it could prove its worth.

Arguably, such a “spidey sense,” no matter how sophisticated, won't cut it for Spiderman. The ability to climb vertical surfaces would be a nice touch, and indeed, researchers from Mark Cutkosky's lab accomplished this by optimizing an adhesive system inspired by geckos (Hawkes et al., 2015). Geckos' uncanny ability to defy gravity resides in spatula-shaped lamellae covering their footpads, which adhere to surfaces through van der Waals forces. A variety of biomimetic materials capturing their properties have been developed; however, the remaining challenge was scaling the amount of adhesive to human weight for safe and uninterrupted climbing. Cutkosky's lab resolved this problem by developing a load-sharing method designed to ensure a uniform distribution of forces across the adhesive, enabling a human to climb a vertical glass surface with a hand-sized area of adhesive.

Camouflage or ability to merge with the surroundings has long been on the wish list of the military industry and consumer product designers. It is in equally high demand in nature and is brought to perfection by cephalopods that can change coloration quickly by altering pigment-containing skin chromatophores. The laboratory of John Rogers



Spiderman (image from iStockphoto).

recently recreated this process by combining several layers of synthetic materials and sensors, mimicking the elements of the cephalopod skin, and achieving autonomous color matching to the background (Yu et al., 2014). Such a system not only has diverse applications, but it also deconstructs the complexity of a natural organ.

Ultimately, biomimetic advances like the ones described above show that harnessing designs honed by evolutionary forces rather than the human creative process may not only result in versatile solutions, but also inspire quests for enhancing human capabilities.

REFERENCES

- Hawkes, E.W., Eason, E.V., Christensen, D.L., and Cutkosky, M.R. (2015). *J.R. Soc. Interface* 12, 20140675.
- Kang, D., Pikhitsa, P.V., Choi, Y.W., Lee, C., Shin, S.S., Piao, L., Park, B., Suh, K.Y., Kim, T.I., and Choi, M. (2014). *Nature* 516, 222–226.
- de Mestral, G. May 1958. Separable fastening device. U.S. patent 3083737 A.
- Yu, C., Li, Y., Zhang, X., Huang, X., Malyarchuk, V., Wang, S., Shi, Y., Gao, L., Su, Y., Zhang, Y., et al. (2014). *Proc. Natl. Acad. Sci. USA* 111, 12998–3003.

Mirna Kvajo

Donald Metcalf (1929–2014)

Once more unto the breach, dear friends, once more,
Or close the wall up with our English dead!
In peace there's nothing so becomes a man
As modest stillness and humility,
But when the blast of war blows in our ears,
Then imitate the action of the tiger:
Stiffen the sinews, summon up the blood.

—William Shakespeare
(King Henry, *Henry V*)

Donald Metcalf—Professor Metcalf to those younger colleagues who were meeting him for the first time and Don to everyone else—was one of the fathers of modern hematology, shaping the field for the 60 years of his working life (1954–2014) at The Walter and Eliza Hall Institute of Medical Research.

Don was born in Mittagong, Australia, a little more than 100 km southwest of Sydney, on February 26, 1929, just prior to the start of the depression. Don's parents were schoolteachers, and his childhood was nomadic, moving with them from school to school in country New South Wales. As one might expect, this environment instilled the self-reliance, resilience, work ethic, and appreciation for education that characterized Don's entire life. Following school, Don entered Sydney University as a medical student in 1946 and completed a research year in the laboratory of Pat De Burgh, working on ectromelia virus. That year of research whetted Don's appetite, and after graduating in 1953, Don was awarded the Carden Fellowship of the Anti-Cancer Council of Victoria and moved to Melbourne in 1954 to establish a cancer research laboratory at The Walter and Eliza Hall Institute of Medical Research, which was at that time led by the

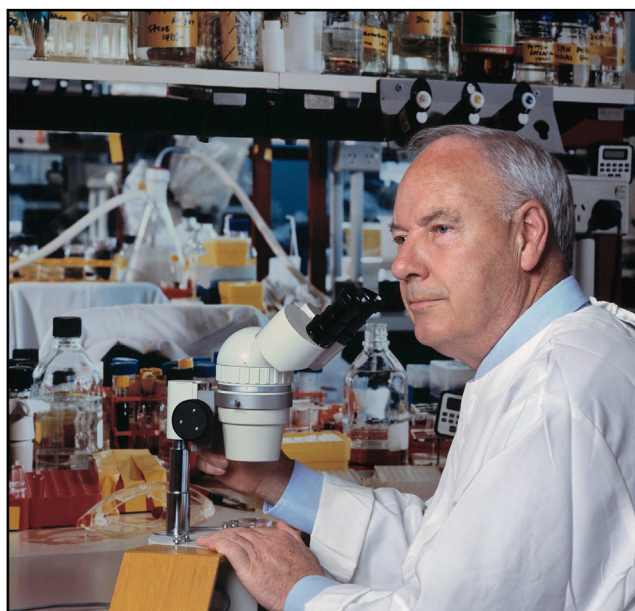
future Nobel Laureate Frank McFarlane Burnet.

In an age when scientific superstars run mega-labs and often move from institution to institution chasing ever more lucrative packages, the opposite was true of Don. Throughout his entire career, he was supported by the same fellowship and worked at the same institution. Despite his burgeoning reputation, he never left the laboratory bench and conducted experiments himself with the support of one or two research assistants and an occasional graduate student. Having studied thymus development and lymphomagenesis for a decade, Don's epiphany came in 1965, when with Ray Bradley at The University of Melbourne, he used semi-solid culture medium containing a crude source of growth factor to grow colonies of granulocytes and macrophages from their precursors in the bone marrow. Don realized immediately that this technique would not only permit him to map the hierarchical relationship between multipotential stem cells, committed progenitors, and their differentiating progeny, but would also allow him to define the hormones that regulated the process. Don worked on nothing else for the next

50 years and, in doing so, made the blood cell system the poster child for understanding tissue renewal.

Don named the hormones that stimulated the growth of blood cells colony-stimulating factors, or CSFs. Unlike classic hormones, which are produced by a single tissue or cell type and act widely, Don found CSFs to be produced by almost every tissue he examined, but in vanishingly small amounts, and to act on a limited set of target cells. This resulted in a huge degree of skepticism in the scientific community; many thought that CSFs were *in vitro* artifacts that were of little importance to the regulation of blood cell production *in vivo*. Skepticism was the spur that Don relished. Recruiting a series of protein chemists and molecular biologists to the collaborative cause, including Richard Stanley, Tony Burgess, Nicos Nicola, Ashley Dunn, and Nick Gough, Don orchestrated a 30-year collaboration that saw the purification to homogeneity of four colony-stimulating factors: granulocyte CSF (G-CSF), macrophage CSF (M-CSF), granulocyte-macrophage CSF (GM-CSF), and multi-CSF, which is now better known as interleukin-3.

The purification of the CSFs led to the cloning of their genes—GM-CSF by Don's team in Melbourne and G-CSF, M-CSF, and IL-3 by others—and ultimately led to the mass production of CSFs, allowing Don to do the experiment about which he had dreamed for 30 years: injecting pure recombinant CSFs into animals and seeing whether blood cell production becomes elevated. The answer was an emphatic yes, and these *in vivo* experiments paved the way for the successful use of G-CSF and GM-CSF in elevating the level of white cells in patients undergoing chemotherapy for cancer to reduce the likelihood of infection due to leukopenia. It was during



Donald Metcalf

these clinical trials that Don and a post-doctoral fellow, Uli Dührsen, discovered that, upon injection of patients with G-CSF, stem cells migrated from the bone marrow into the peripheral blood. This observation rapidly led to the use of G-CSF as an agent to mobilize stem cells, meaning that they could be obtained by simple blood collection rather than the more complicated and painful bone marrow harvest. This revolutionized bone marrow transplantation, making it safer, more effective, and more broadly applicable. It is estimated that more than 20 million people have benefitted from Don's discoveries, including the Spanish tenor José Carreras, who was an early recipient of the therapy.

Don's profound discoveries led to countless distinguished lectureships, awards, and prizes, including Fellowship of the Australian Academy of Science, The Royal Society and the US National Academy of Sciences, The Bristol-Myers Award for Distinguished Achievement in Cancer Research, The Robert Koch Prize, The Armand Hammer Prize for Cancer Research, The General Motors Cancer Foundation Sloan Prize, The Lasker-DeBakey Clinical Medical Research Award, and a Lifetime Achievement Award from the American Association for Cancer Research. Although his profound discoveries and many honors epitomize the scientist, they don't paint a complete picture of the man.

Don detested spin—he wanted his data to speak for itself and was scathing of researchers he termed “prancers” or “strutters,” who were more showmen than sci-

entists. At conferences and in seminars, he would ask the incisive question or, if required, deliver the acerbic assessment—the more famous the researcher, the more pointed and colorful the criticism if the talk was not up to par. He was happy to send his papers to specialty hematology journals, and unless badgered by ambitious junior colleagues, Don eschewed what he called “fancy” journals, which he thought required sensationalization of the story.

Don was comfortable working in Melbourne, which for those in Boston, San Francisco, London, or Paris seemed like the end of the Earth. This tyranny of distance was actually a great source of comfort to Don, who could follow his own scientific compass without being blown off course by trends or fads. Don was deeply suspicious of researchers who, on achieving a level of success, retreated out of the laboratory into the safety of their office. He was a man who led from the experimental front for his whole life, which is why he so liked the “Once more unto the breach” speech of *Henry V* and titled his autobiography *Summon Up The Blood*. He worked five and a half days a week for most of the last 50 years, with his only concession to age being an occasional Saturday off and leaving work a little early to avoid the traffic in the evening.

Other than a bad back, which was no doubt exacerbated by endless hours working at the microscope, Don remained in remarkable health until the middle of last year. Feeling “off color,” Don and his beloved wife of 60 years, Josephine,

took a vacation, which he hoped would reinvigorate him for another decade of discovery. Returning in August feeling worse, Don suspected the worst and was quickly diagnosed with metastatic pancreatic cancer. Knowing the chance of cure was slim, Don had two concerns: to finish off the experiments awaiting his return from vacation and to spend as much time as possible with his family. But how to do both? The answer was to have his microscope moved from his laboratory to his dining room table at home and to do experiments when his chemotherapy schedule allowed. Don carried out his last experiment in November, when his health began to decline precipitously. He died on December 15, 2014, surrounded by his wife Jo and their four daughters and their families. Don was an experimentalist and a family man to the end. He would have wanted it no other way.

Douglas J. Hilton,^{1,2,*}
Nicos A. Nicola,^{1,2}
Warren S. Alexander,^{1,2}
Andrew W. Roberts,^{1,2,3}
and Ashley R. Dunn⁴

¹The Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville 3052, Victoria, Australia

²Department of Medical Biology, The University Of Melbourne, Parkville 3010, Victoria, Australia

³The Royal Melbourne Hospital, Parkville 3050, Victoria, Australia

⁴Department of Surgery, The University Of Melbourne, Parkville 3010, Victoria, Australia

*Correspondence: hilton@wehi.edu.au
<http://dx.doi.org/10.1016/j.cell.2015.01.017>

The Metabolic Milieu of Metastases

William J. Sullivan¹ and Heather R. Christofk^{1,2,3,*}

¹Department of Molecular and Medical Pharmacology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA

²Jonsson Comprehensive Cancer Center, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA

³Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, University of California, Los Angeles, Los Angeles, CA 90095, USA

*Correspondence: hchristofk@mednet.ucla.edu
<http://dx.doi.org/10.1016/j.cell.2015.01.023>

To colonize the liver, colon cancer metastases must overcome hypoxia and other metabolic stress. Loo et al. now show that metastatic cells achieve this by decreasing miR-483 and miR-551a expression, which derepresses creatine kinase expression and allows energy to be captured from extracellular ATP through generation and import of phosphocreatine.

It has long been appreciated that cancer cells exhibit altered patterns of metabolism. The century-old observation of aerobic glycolysis—or, the Warburg Effect—has evolved into a more nuanced and context-specific understanding of how cancer cells reprogram their metabolism to meet the biosynthetic demands of rapid proliferation and overcome metabolic stress imposed by the microenvironment (Hanahan and Weinberg, 2011). In this issue of *Cell*, in an elegant series of experiments, Loo et al. (2015) now demonstrate that downregulation of specific microRNAs (miRNAs) allows metastatic colon cancer cells to adapt metabolically to the harsh conditions in the liver by using a secreted liver metabolite to scavenge energy from the extracellular environment.

Several lines of evidence have suggested key roles for miRNAs in cancer progression and metabolism. Broad dysregulation of miRNA expression accompanies tumorigenesis (Volinia et al., 2006), while loss of specific miRNAs is associated with promotion of metastasis both experimentally and clinically (Tava-zoie et al., 2008). miRNAs also regulate key enzymes important for maintaining the cancer metabolic program. For example, MYC enhances cancer glutamine metabolism in part by suppressing miR-23a/b, leading to upregulation of the target glutaminase (Gao et al., 2009), and mTOR activation promotes cancer glycolysis in part by suppressing miR-143 and upregulating its target, hexokinase 2 (Fang et al., 2012). Using two

independent in vivo selection techniques in mice, Loo et al. demonstrate that silencing of miR-483 and miR-551a, both of which target creatine kinase, brain-type (CKB), promotes metastasis of colon cancer cells to the liver. But how does derepression of CKB expression enhance metastatic colonization?

The liver is considerably hypoxic and may have uneven levels of glucose in the interstitial space due to competition from neighboring hepatocytes. Because oxygen and glucose are key substrates for oxidative and glycolytic metabolism, colon cancer cells that metastasize to the liver experience metabolic stress. To overcome this stress, they capitalize on another metabolic feature of the liver—synthesis and secretion of creatine by hepatocytes. By suppressing miR-483 and miR-551a, colon cancer cells upregulate expression of their common target, CKB, and secrete it into the extracellular space, where it catalyzes formation of phosphocreatine from creatine and ATP. Phosphocreatine is then imported back into colon cancer cells through the transporter SLC6A8, where it is used to regenerate ATP needed for cellular functions (Figure 1). Thus, by downregulating miR-483 and miR-551a, metastatic colon cancer cells are able to bypass intracellular glycolytic and oxidative metabolism by activating an alternative metabolic pathway, whereby they scavenge energy from the extracellular environment in the form of phosphocreatine and shuttle it into the cell.

Emerging evidence suggests that there may be a high degree of organ specificity in selection for metabolic traits enabling successful metastatic dissemination. While Loo et al. define a metabolic selection mechanism for colon cancer colonization of the liver, previous work has shown that fatty acids secreted by adipocytes in the omentum fuel growth of ovarian cancer metastases expressing higher levels of fatty acid-binding protein 4—a metabolic explanation for the predilection of ovarian cancer cells to metastasize to the omentum compared to neighboring tissues and organs (Nieman et al., 2011). Other sites of metastasis may also have unique macrometabolic properties that enhance metastatic capacity of cancer cells with complementary metabolic proclivities. For instance, two common sites of metastasis, the brain and the lungs (Nguyen et al., 2009), have high basal levels of glucose and oxygen perfusion, respectively, which may contribute to the high degree of metastatic colonization of these organs compared with others.

Notably, loss of miR-483 and miR-551a and increased expression of CKB, as well as SLC6A8, are observed in clinical samples of colon cancer liver metastases compared to primary colon tumors, suggesting a possible therapeutic opportunity. Indeed, the authors show that viral rescue of both miR-483 and miR-551a, as well as pharmacological inhibition of CKB with cyclocreatine, substantially decrease the extent of liver metastasis in a mouse model of metastatic colon

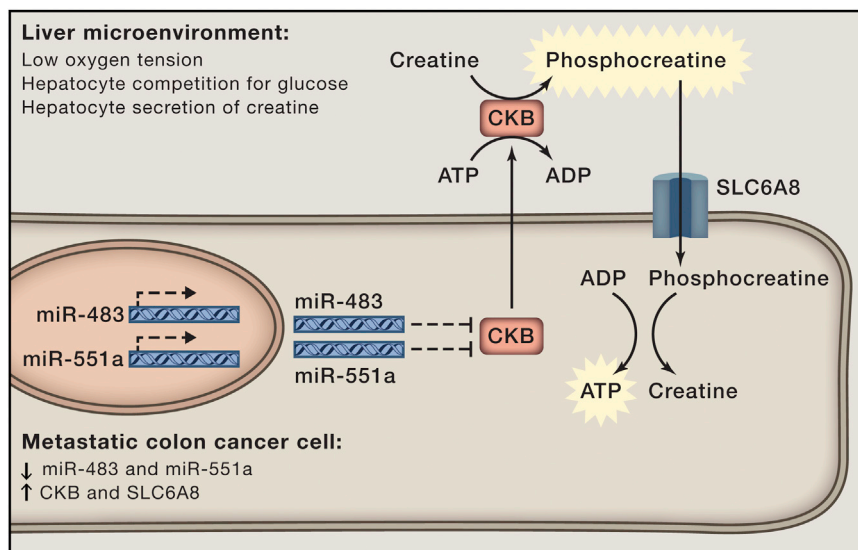


Figure 1. Phosphocreatine Fuels Metastatic Colonization of the Liver

Reduced miR-483 and miR-551a levels in metastatic colon cancer cells liberates expression of their mutual target, brain-type creatine kinase (CKB). CKB secreted by the metastatic cells in the liver phosphorylates extracellular creatine produced by hepatocytes using extracellular ATP to generate phosphocreatine. Extracellular phosphocreatine is imported into metastatic colon cancer cells by the transporter SLC6A8. Phosphocreatine is then used to regenerate the ATP needed to sustain the cell. This alternative metabolic pathway using extracellular phosphocreatine enables metastatic cells to survive metabolic stress imposed by the liver microenvironment during initial organ colonization, before tumor vascularization delivers canonical metabolic substrates.

cancer. Importantly, the reliance on phosphocreatine as an energy source for liver metastases may not be limited to colon cancer. Loo et al. show that pancreatic cancer cells, which also metastasize to the liver, suffer a similar decrease in metastatic capacity when CKB levels are depleted. Little has been reported on CKB levels in cancer, but it will be interesting to assess whether CKB levels correlate with a propensity to metastasize to the liver in other cancer subtypes.

The findings by Loo et al. suggest that targeting phosphocreatine metabolism may reduce liver metastasis in colon cancer patients. However, to successfully translate these findings to the clinic, it will be important to determine whether reliance of liver metastases on phosphocreatine metabolism changes at different stages of disease. Viral rescue of miR-483 and miR-551a and pharmacological inhibition of CKB reduce metastatic tumor burden when administered within 24 hr of

injecting mice with colon cancer cells, suggesting that targeting phosphocreatine metabolism could be of clinical utility for patients with early disseminated disease to prevent development of liver metastases. But for patients with detectable secondary lesions, the value of targeting phosphocreatine metabolism remains unclear. Phosphocreatine metabolism may be most important for cancer cells to combat metabolic stress upon initial seeding of the liver, before expansion into a secondary tumor, when tumor-induced vasculature can deliver canonical metabolic substrates. Bolstering the idea that initial organ colonization and subsequent tumor expansion require different metabolic programs, circulating breast cancer cells were recently shown to exhibit increased oxidative metabolism compared to the primary tumor and secondary tumor outgrowths in the lungs (LeBleu et al., 2014). Additionally, liver metastases in colorectal cancer patients

exhibit glycolytic metabolism, according to PET studies using the glucose analog ^{18}F -fluorodeoxyglucose (Maffione et al., 2015).

Together, these data suggest that metastatic cells may require a metabolic program focused on energy generation to survive metabolic stress upon initial colonization of a site, and may switch to a program using canonical metabolic substrates for growth post-vascularization. Understanding the dynamics of metabolic transitions between initial cancer dissemination, liver colonization, and metastatic outgrowth will aid selection of colon cancer patients who will benefit most from approaches to target phosphocreatine metabolism.

REFERENCES

- Fang, R., Xiao, T., Fang, Z., Sun, Y., Li, F., Gao, Y., Feng, Y., Li, L., Wang, Y., Liu, X., et al. (2012). *J. Biol. Chem.* 287, 23227–23235.
- Gao, P., Tchernyshyov, I., Chang, T.C., Lee, Y.S., Kita, K., Ochi, T., Zeller, K.I., De Marzo, A.M., Van Eyk, J.E., Mendell, J.T., and Dang, C.V. (2009). *Nature* 458, 762–765.
- Hanahan, D., and Weinberg, R.A. (2011). *Cell* 144, 646–674.
- LeBleu, V.S., O'Connell, J.T., Gonzalez Herrera, K.N., Wikman, H., Pantel, K., Haigis, M.C., de Carvalho, F.M., Damascena, A., Domingos Chinen, L.T., Rocha, R.M., et al. (2014). *Nat. Cell Biol.* 16, 992–1003, 1–15.
- Loo, J.M., Scherl, A., Nguyen, A., Man, F.M., Weinberg, E., Zeng, Z., Saltz, L., Paty, P.B., and Tavazoie, S.F. (2015). *Cell* 160, this issue, 393–406.
- Maffione, A.M., Lopci, E., Bluemel, C., Giammarile, F., Herrmann, K., and Rubello, D. (2015). *Eur. J. Nucl. Med. Mol. Imaging* 42, 152–163.
- Nguyen, D.X., Bos, P.D., and Massagué, J. (2009). *Nat. Rev. Cancer* 9, 274–284.
- Nieman, K.M., Kenny, H.A., Penicka, C.V., Ladanyi, A., Buell-Gutbrod, R., Zillhardt, M.R., Romero, I.L., Carey, M.S., Mills, G.B., Hotamistligil, G.S., et al. (2011). *Nat. Med.* 17, 1498–1503.
- Tavazoie, S.F., Alarcón, C., Oskarsson, T., Padua, D., Wang, Q., Bos, P.D., Gerald, W.L., and Massagué, J. (2008). *Nature* 451, 147–152.
- Volinia, S., Calin, G.A., Liu, C.G., Ambs, S., Cimmino, A., Petrocca, F., Visone, R., Iorio, M., Roldo, C., Ferracin, M., et al. (2006). *Proc. Natl. Acad. Sci. USA* 103, 2257–2261.

Growing Old with Myc

Patrick A. Carroll¹ and Robert N. Eisenman^{1,*}

¹Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA

*Correspondence: eisenman@fhcrc.org

<http://dx.doi.org/10.1016/j.cell.2015.01.018>

The *Myc* proto-oncogene has been intensively studied in tumorigenesis and development. A new paper in *Cell* reports the role of *Myc* as a determinant of mammalian longevity. *Myc* heterozygous mice exhibit extended lifespans resulting from alterations in multiple cellular processes distinct from those observed in other longevity models.

Myc ranks among the most exhaustively studied genes in the vertebrate genome. The intense interest stems from its essential and multiple roles in development and its critical functions in the etiology of most, if not all, cancers. The *Myc* protein is thought to function predominantly as a transcriptional regulator influencing the expression of thousands of genes and broadly promoting the fundamental cellular processes of growth, metabolism, proliferation, differentiation, and death in a context-dependent manner (for reviews, see [Dang and Eisenman, 2014](#)). While *Myc* has been intensively studied at the cellular level, relatively little is known concerning its roles in organismal physiology. In this issue of *Cell*, Sedivy and coworkers ([Hofmann et al., 2015](#)) provide compelling evidence that mammalian longevity and healthspan are linked in multiple ways to the abundance of the *Myc* protein.

By comparing wild-type mice (*Myc*^{+/+}) with mice expressing roughly half the amount of *Myc* mRNA and protein (*Myc*^{+/-}), Hofmann et al. found a 15% increase in median lifespan (combined for both sexes) caused by *Myc* hypomorphism. Importantly, the attenuated mortality is manifested across all ages, indicating that the impact of lowered *Myc* levels is not restricted to a specific period of time. This augmentation in lifespan appears to be accompanied by a diverse group of phenotypes that are indicative of an increased healthspan and correlate with other long-lived mouse models, such as calorie restricted (CR) and Ames dwarf mice ([Gems and Partridge, 2013](#)). These phenotypes include decreases in body size, serum IGF1 levels, immunosenescence, spontaneous tumor progression, osteoporosis, and serum and liver cholesterol content, as well as enhanced

metabolic rate and neuromuscular performance, all of which are characteristic of young animals and are observed in older *Myc* hypomorphic mice. While phenotypic similarities between *Myc*^{+/-} and other lifespan-extending conditions are apparent, there are also some significant differences. For example, in the *Myc* hypomorphic mice, no changes are detected in body temperature, adiposity, or fertility, while all three are changed in other mouse models of increased longevity. Moreover, meta-analyses of gene expression profiles derived from mice with extended lifespan due to CR or metformin treatment reveal only few changes in common with those from the *Myc* hypomorphic mice. Together, these data suggest that lower *Myc* levels promote lifespan extension through a unique constellation of molecular events. In addition, as the authors point out, major consequences of decreased *Myc* may be indirect and not necessarily apparent at the transcriptional level. For instance, diminished rates of translation or suppression of mTOR activity have been linked to lifespan extension in flies, yeast, and nematodes ([Gems and Partridge, 2013](#)).

Previous studies in *Drosophila melanogaster* have implicated members of the *Myc* transcriptional network in the regulation of aging. Loss-of-function alleles of the *Drosophila* *Myc* antagonist, dMnt, increase cell size and body weight while decreasing fly lifespan ([Loo et al., 2005](#)). More recently, *Drosophila* *Myc* has been shown to act as a rheostat for longevity, as its overexpression diminishes and heterozygous deletion augments longevity in flies ([Greer et al., 2013](#)). These phenotypes appear to be independent of apoptosis and correlated with

dMyc-induced genomic instability. In the *Myc*^{+/-} mice, while apoptosis also appears unaffected, markers of DNA damage and senescence are unchanged and cells derived from these mice are equally sensitive to toxic chemicals compared to the wild-type controls, suggesting that the increased longevity and healthspan of the *Myc* hypomorphic mice are not likely to be due to resistance to genomic instability, oxidative stress, or cellular insults. However, it is yet to be determined whether the increased healthspan and lifespan could be due to a decreased frequency of insult.

Potential candidates for driving the longevity phenotype include diminished serum IGF1 levels and increased metabolic rate, which are shared by *Myc*^{+/-} mice and the somatotrophic mutant mouse strains, such as the Ames dwarf mice, and the CR mice ([Lee and Longo, 2011](#)). Importantly, IGF1 regulates mTOR activity via the PI3K-AKT pathway, and *Myc*^{+/-} mice exhibit decreased IGF1 and mTOR activity, as well as decreased rRNA production and translation rate. In summary, these mice appear to display better proteostasis. Consistent with these findings, treatment with the mTOR inhibitor rapamycin or haploinsufficiency of mTOR components or targets has been shown to lead to increased lifespan ([Johnson et al., 2013](#)).

Alterations in metabolism, as evidenced by enhanced oxygen-consumption and mitochondrial content and function, have also been associated with longevity. In *Drosophila*, expression of the PGC-1 transcriptional coactivator in intestinal stem cells (ISCs) results in stimulation of mitochondrial biogenesis and oxidative phosphorylation, which limits ISC proliferation and misdifferentiation,

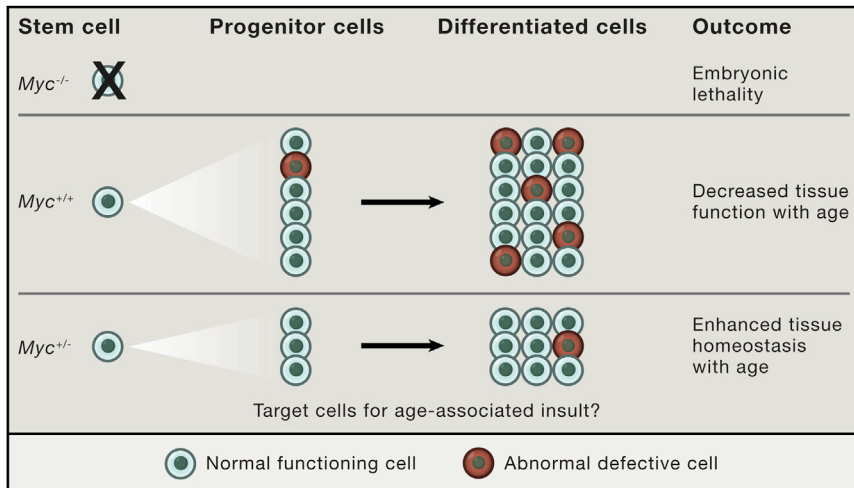


Figure 1. Myc Levels Dictate Tissue Homeostasis, Organismal Size, and Lifespan

Mice lacking the *Myc* gene (*Myc*^{-/-}) exhibit embryonic lethality, whereas compared with wild-type (*Myc*^{+/+}) controls, *Myc*^{+/-} mice exhibit enhanced healthspan and lifespan. The lower abundance of Myc protein in *Myc* hypomorphic mice results in a reduction in the number of proliferative progenitors and in the organ size, possibly by suppressing stem cell exhaustion and limiting the number of defective or misdifferentiated cells within the tissue, thereby maintaining tissue homeostasis.

attenuates an age-dependent loss of tissue homeostasis, and promotes maintenance of a functional stem cell population characteristic of younger organisms (Rera et al., 2011). Thus, metabolic changes may help to maintain tissue integrity by imposing a differentiation pace that does not deplete somatic quiescent stem cells. It has been well established that Myc exerts a profound effect on metabolic activity and mitochondrial function and regulates the size of stem, progenitor, and differentiated cell compartments (Dang and Eisenman, 2014; Freije et al., 2014; Laurenti et al., 2008). Indeed, Hofmann et al. detect a higher ratio of quiescent long-term to proliferative short-term hematopoietic stem cells in *Myc* hypomorphic mice. This finding suggests that the loss of one *Myc* allele might be sufficient to limit the maximal abundance of

Myc protein attained during stem and progenitor cell proliferation and differentiation, thereby dampening stem cell depletion, curtailing progenitor cell overgrowth, and positively affecting homeostasis of multiple cell lineages (Figure 1).

Like most important findings, the work of Hofmann et al. on *Myc*'s role in longevity raises more questions than it answers. One outstanding question concerns the tissues that are most directly relevant to longevity in the *Myc* hypomorphic mice. Is the longevity phenotype due to changes in multiple tissue types, or does one tissue predominate by triggering a cascade of events that ultimately extends both lifespan and healthspan? In particular, do nervous system-specific changes that lead to increased neuromuscular fitness and metabolic alterations contribute to better healthspan? Is the observed decrease in translation

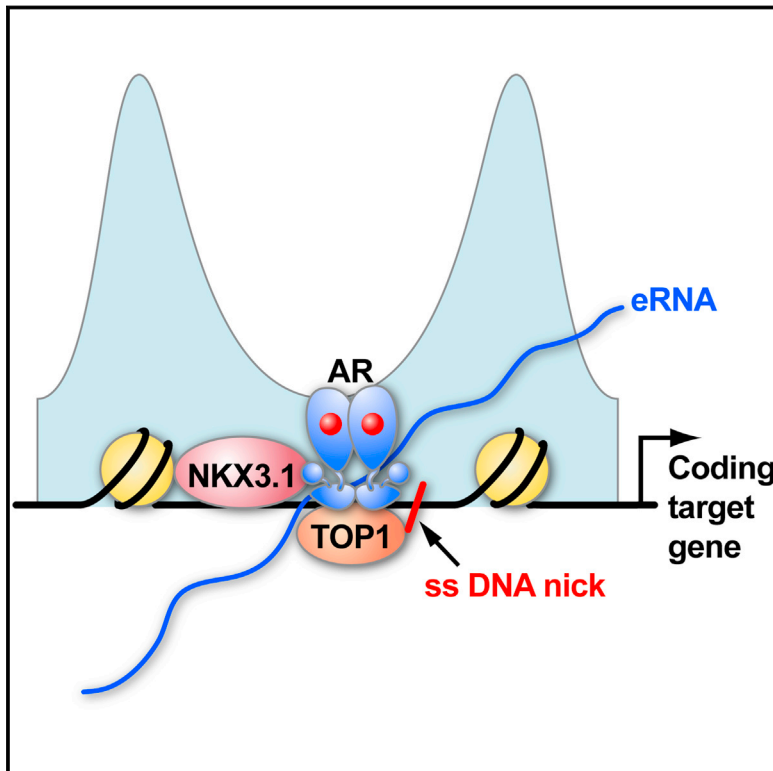
rate sufficient to account for the decrease in cholesterol biosynthetic enzymes and the lower production of IGF1 in the liver and serum? Could enhanced proteostasis resulting from diminished translational stress affect every tissue where protein quality control is important? Another interesting question is whether other members of the extended *Myc* network also play a role in regulating aging. It is evident that the more we know about even such a well-studied gene as *Myc*, the more we have to learn.

REFERENCES

- Dang C.V. and Eisenman R.N., eds. (2014). *Myc and the Pathway to Cancer* (Cold Spring Harbor, N.Y.: Cold Spring Harbor Press).
- Freije, A., Molinuevo, R., Ceballos, L., Cagigas, M., Alonso-Lecue, P., Rodriguez, R., Menendez, P., Aberdam, D., De Diego, E., and Gandarillas, A. (2014). *Cell Rep.* 9, 1349–1360.
- Gems, D., and Partridge, L. (2013). *Annu. Rev. Physiol.* 75, 621–644.
- Greer, C., Lee, M., Westerhof, M., Milholland, B., Spokony, R., Vijg, J., and Secombe, J. (2013). *PLoS ONE* 8, e74641.
- Hofmann, J.W., Zhao, X., De Cecco, M., Peterson, A.L., Pagliaroli, L., Manivannan, J., Hubbard, G.B., Ikeno, Y., Zhang, Y., Feng, B., et al. (2015). *Cell* 160, this issue, 477–488.
- Johnson, S.C., Rabinovitch, P.S., and Kaeblerlein, M. (2013). *Nature* 493, 338–345.
- Laurenti, E., Varnum-Finney, B., Wilson, A., Ferrero, I., Blanco-Bose, W.E., Ehninger, A., Knoepfler, P.S., Cheng, P.-F., MacDonald, H.R., Eisenman, R.N., et al. (2008). *Cell Stem Cell* 3, 611–624.
- Lee, C., and Longo, V.D. (2011). *Oncogene* 30, 3305–3316.
- Loo, L.W., Secombe, J., Little, J.T., Carlos, L.S., Yost, C., Cheng, P.F., Flynn, E.M., Edgar, B.A., and Eisenman, R.N. (2005). *Mol. Cell. Biol.* 25, 7078–7091.
- Rera, M., Bahadorani, S., Cho, J., Koehler, C.L., Ulgherait, M., Hur, J.H., Ansari, W.S., Lo, T., Jr., Jones, D.L., and Walker, D.W. (2011). *Cell Metab.* 14, 623–634.

Ligand-Dependent Enhancer Activation Regulated by Topoisomerase-I Activity

Graphical Abstract



Authors

Janusz Puc, Piotr Kozbial, ..., Aneel K. Aggarwal, Michael G. Rosenfeld

Correspondence

mrosenfeld@ucsd.edu

In Brief

The requirement for Topoisomerase 1 nicking function in enhancer activation reveals an unexpected connection between transcription and DNA Damage Repair response and suggests that DNA nicking is necessary to relieve torsional stress and promote eRNA synthesis.

Highlights

- DNA topoisomerase I (TOP1) is recruited to AR-bound functional enhancers
- TOP1 nicking function is required for enhancer activation
- Recruitment of the MRN complex including MRE11 modulates enhancer activation
- Temporarily late recruitment of DDR machinery is observed on AR-bound enhancers

Accession Numbers

GSE63202



Ligand-Dependent Enhancer Activation Regulated by Topoisomerase-I Activity

Janusz Puc,¹ Piotr Kozbial,¹ Wenbo Li,¹ Yuliang Tan,¹ Zhijie Liu,¹ Tom Suter,^{1,2} Kenneth A. Ohgi,¹ Jie Zhang,¹ Aneel K. Aggarwal,³ and Michael G. Rosenfeld^{1,*}

¹Howard Hughes Medical Institute, Department of Medicine, School of Medicine, University of California, San Diego, La Jolla, CA 92093-0648, USA

²Division of Biological Sciences, Department of Molecular Biology, University of California, San Diego, La Jolla, CA 92093-0648, USA

³Department of Structural and Chemical Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

*Correspondence: mrosenfeld@ucsd.edu

<http://dx.doi.org/10.1016/j.cell.2014.12.023>

SUMMARY

The discovery that enhancers are regulated transcription units, encoding eRNAs, has raised new questions about the mechanisms of their activation. Here, we report an unexpected molecular mechanism that underlies ligand-dependent enhancer activation, based on DNA nicking to relieve torsional stress from eRNA synthesis. Using dihydrotestosterone (DHT)-induced binding of androgen receptor (AR) to prostate cancer cell enhancers as a model, we show rapid recruitment, within minutes, of DNA topoisomerase I (TOP1) to a large cohort of AR-regulated enhancers. Furthermore, we show that the DNA nicking activity of TOP1 is a prerequisite for robust eRNA synthesis and enhancer activation and is kinetically accompanied by the recruitment of ATR and the MRN complex, followed by additional components of DNA damage repair machinery to the AR-regulated enhancers. Together, our studies reveal a linkage between eRNA synthesis and ligand-dependent TOP1-mediated nicking—a strategy exerting quantitative effects on eRNA expression in regulating AR-bound enhancer-dependent transcriptional programs.

INTRODUCTION

Research over the past few years, supported by data from GRO sequencing (GRO-seq) analysis and the ENCODE project, has revealed that most developmental and regulatory transcriptional regulation programs are controlled by an extensive enhancer network (Kim et al., 2010; Shlyueva et al., 2014), with each cell type estimated to harbor 70,000–100,000 enhancers, located upstream and downstream of coding target gene promoters (Pennacchio et al., 2013). Enhancer signatures include monomethylated H3K4 (H3K4me1) and H3K27-acetylated histones (Kim et al., 2010; Li et al., 2013a; Wang et al., 2011). These enhancers are usually characterized by a nucleosome-depleted core region where many of the cooperating transcription factors bind (Andersson et al., 2014; Hah et al., 2013; Kaikkonen et al., 2013; Lai

et al., 2013; Lam et al., 2013; Li et al., 2013a; Melgar et al., 2011; Melo et al., 2013; Mousavi et al., 2013). Most surprisingly, enhancers are also transcription units, wherein their effect on target coding genes correlates with the transcription of the lncRNAs, referred to as eRNAs (Andersson et al., 2014; De Santa et al., 2010; Hah et al., 2013; Kaikkonen et al., 2013; Kim et al., 2010; Lai et al., 2013; Lam et al., 2013; Li et al., 2013a; Melgar et al., 2011; Melo et al., 2013; Mousavi et al., 2013) adding a new layer of regulation to the fundamental mechanisms underlying enhancer action (Lam et al., 2014; Natoli and Andrau, 2012).

The current prevailing belief, based on chromosome capture assays, where looping constraints are inferred from interaction frequencies between a point of interest and distal loci of the genome, is that the main mechanism by which enhancers affect their target gene expression is through chromatin looping. eRNA transcripts seem to be functionally important by contributing to the stabilization of juxtaposed enhancer-target gene promoter loops to allow for optimal gene expression (Lai et al., 2013; Li et al., 2013a). However, both eRNA synthesis and nucleosome depletion are potential sources of topological strain on enhancers that can possibly hinder transcription. The movement and rotation of RNA polymerase complex (RNAP) along DNA template during the process of RNA synthesis (Liu and Wang, 1987) can generate positive supercoils in front of the advancing RNAP, and negative supercoils behind it (Darzacq et al., 2007; Kouzine et al., 2013; Kouzine and Levens, 2007; Liu and Wang, 1987). Because RNA polymerase is a powerful torsional motor, it can alter DNA topology by creating DNA supercoils, which can propagate and affect transcription elongation (Ma and Wang, 2014). Although negative supercoiling can initially facilitate transcription initiation, either by helping RNAP to form an open complex or by helping to recruit transcription factors (Ma and Wang, 2014), it can subsequently lead to the generation of R-loops resulting from hybridization of nascent RNA to the DNA strand that is being transcribed, which, in turn, can impede transcriptional elongation (El Hage et al., 2010). Positive or overwound supercoiling can prevent transcription initiation and greatly diminish mRNA synthesis (Ma and Wang, 2014). Moreover, the very depletion of histones from the core region of enhancers releases unconstrained negative supercoils, which can impede transcription factor binding. One mechanism that resolves the undesirable effects of excessive supercoiling employs DNA topoisomerases, including topoisomerase I (TOP1). TOP1 can relax both negative and positive

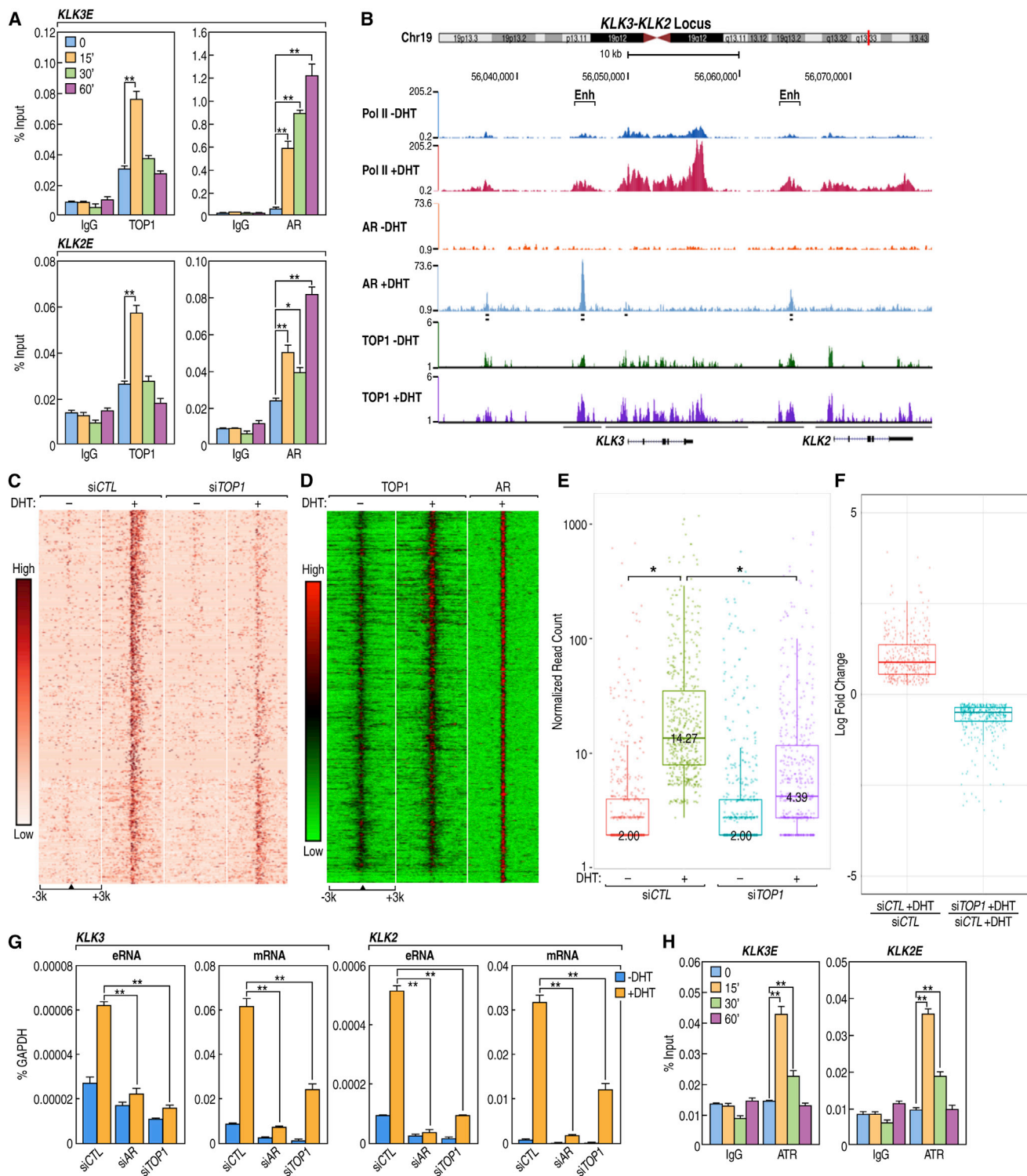


Figure 1. TOP1 Occupies AR Enhancers and Affects the Transcriptional Program of the Prostate Cancer Cell Line LNCaP

(A) Recruitment of AR and TOP1 to the *KLK3* and *KLK2* enhancers. The highest TOP1 binding is detected at 15 min DHT treatment. Data points show mean \pm SD (n = 3), *p < 0.05, **p < 0.01.

(B) The UCSC genome browser screenshot of the *KLK3-CLK2* locus showing the occupancy of p-S5-RNA Pol II (Pol II), AR, and TOP1 (all tested with and without DHT treatment).

(legend continued on next page)

supercoils by transient single-strand breaks for the passage of individual DNA strands through one another, followed by the rejoining of the phosphodiester backbone of DNA (Pedersen et al., 2012; Pommier et al., 2006).

Although TOP1 activity is well established in DNA replication, its potential functionality in enhancer activation and transcriptional initiation remains unclear. Most of the experiments hitherto examining the role of TOP1 in transcription have been limited to artificial promoter model systems, which, if anything, have argued that TOP1 DNA nicking activity is not involved in transcriptional activation in such in vitro systems (Kretzschmar et al., 1993; Merino et al., 1993; Shykind et al., 1997).

However, the utilization of a nicking strategy for transcriptional initiation and enhancer-regulated events would be in concert with the elegant explication of the molecular mechanisms underlying the expression of bacteriophage T4 late genes, with the participation of DNA-mounted activator of transcription, gp45 and RNAP-bound gp33. Here, a nick in the strands of the DNA and the actions of an exonuclease are required, with the DNA template single-strand nicks being essential for transcriptional activation and the nicked-DNA gp45-loading site located upstream or downstream of its target site (Herendeen et al., 1992). Also, in human cells, artificially generated nicks (but not double-strand DNA breaks) have recently been found to be associated with transcription (Davis and Maizels, 2014). Together, these and other experiments in prokaryotes and eukaryotes suggest an intriguing link between DNA nicking and transcription, but the mechanism and the factors involved remain largely unknown.

Here, we describe a molecular mechanism that operates at functional androgen-regulated enhancers and identify DNA topoisomerase I as a critical DNA-nicking enzyme involved in the process of cell-specific, ligand-driven enhancer activation. Recruitment of TOP1 to these AR-bound enhancers is of functional consequence as knockdown of the enzyme in the prostate cancer cells results in inhibition of DHT-regulated eRNA and many coding gene transcriptional targets. Additionally, we provide evidence that recruitment of a significant repertoire of DNA damage response machinery occurs on these functional enhancers, potentially to prevent undesirable effects of persistent DNA damage.

RESULTS

TOP1 Recruitment to AR-Regulated Enhancers Affects eRNA and Coding Gene Expression

To further investigate the mechanism of enhancer activation in ligand-regulated transcription, we employed an early prostate

adenocarcinoma cell line, LNCaP, the growth of which is androgen-dependent (Horoszewicz et al., 1980). The cell line is exquisitely sensitive to androgen stimulation and arrests in the G1 phase of the cell cycle upon steroid depletion, despite the presence of peptide growth factors (Figure S1A). Regulation of cyclin D expression and concomitant CDK4 activity represents one mechanism by which androgen impinges on the cell cycle to govern proliferation (Knudsen et al., 1998).

To investigate whether TOP1 played a role in ligand-regulated transcription, we undertook to examine the possible recruitment of TOP1 to enhancers, finding that it was recruited to several AR enhancers early in response to androgen (5 α -dihydrotestosterone [DHT]) treatment in the ligand-dependent LNCaP prostate cancer cell line (Figures 1A and S1B). These data prompted us to study genome-wide localization of this protein by performing chromatin immunoprecipitation coupled with next-generation sequencing (ChIP-seq). Because exhaustive efforts to identify a TOP1 antibody suitable for ChIP-seq proved unsuccessful, we generated a stable LNCaP cell line with inducible biotinylated TOP1 expression. We observed that TOP1 recruitment in response to DHT, generated enriched regions of a range of sizes (Figure 1B), as opposed to point sources, as found for factors such as the androgen receptor, or broad sources, such as observed for the H3K36me3 histone mark (Sims et al., 2014). Consistent with the observation that enhancers represent regulated transcription units, we noticed a hormone-dependent increase in RNA Pol II (phospho-Ser5) occupancy predominately at these enhancers (Figures 1B and S1C). As expected, we also observed increased TOP1 occupancy over promoters and gene bodies of the representative DHT-induced genes (e.g., *KLK3* and *KLK2*), consistent with the possibility that TOP1 might be involved in both enhancer activation and transcriptional elongation events.

Preliminary analysis demonstrated that TOP1 binding overlapped, in particular, with that of liganded androgen receptor at enhancers (Figures 1B and S1C). Genome-wide analysis revealed 6,545 putative “AR-bound enhancer” sites based on the criterion of an AR-bound locus marked with H3K4me1 and H3K27Ac and more than 1 kb away (in either direction) from the promoter of annotated genes of which 96% bound TOP1, with 3,921 (60%) exhibiting a DHT-stimulated increase in TOP1 binding (Figure S1D).

To assess eRNAs induced by DHT, we took advantage of technological advances that permit mapping of the position, amount, and orientation of transcriptionally engaged RNA polymerase II on a genome-wide scale (Core et al., 2008). GRO-seq analysis (Core et al., 2008) of serum-starved LNCaP cells treated for 1 hr with DHT identified 644 putative AR

(C) GRO-seq analysis of the effect of *TOP1* knockdown on nascent RNA levels shown as a heatmap for 579 enhancers (out of 644, which were upregulated by DHT treatment) with the most affected AR enhancers at the top.

(D) Heatmap showing DHT-induced TOP1 sequencing tags density increase around 644 AR-enhancer binding sites (centered on AR).

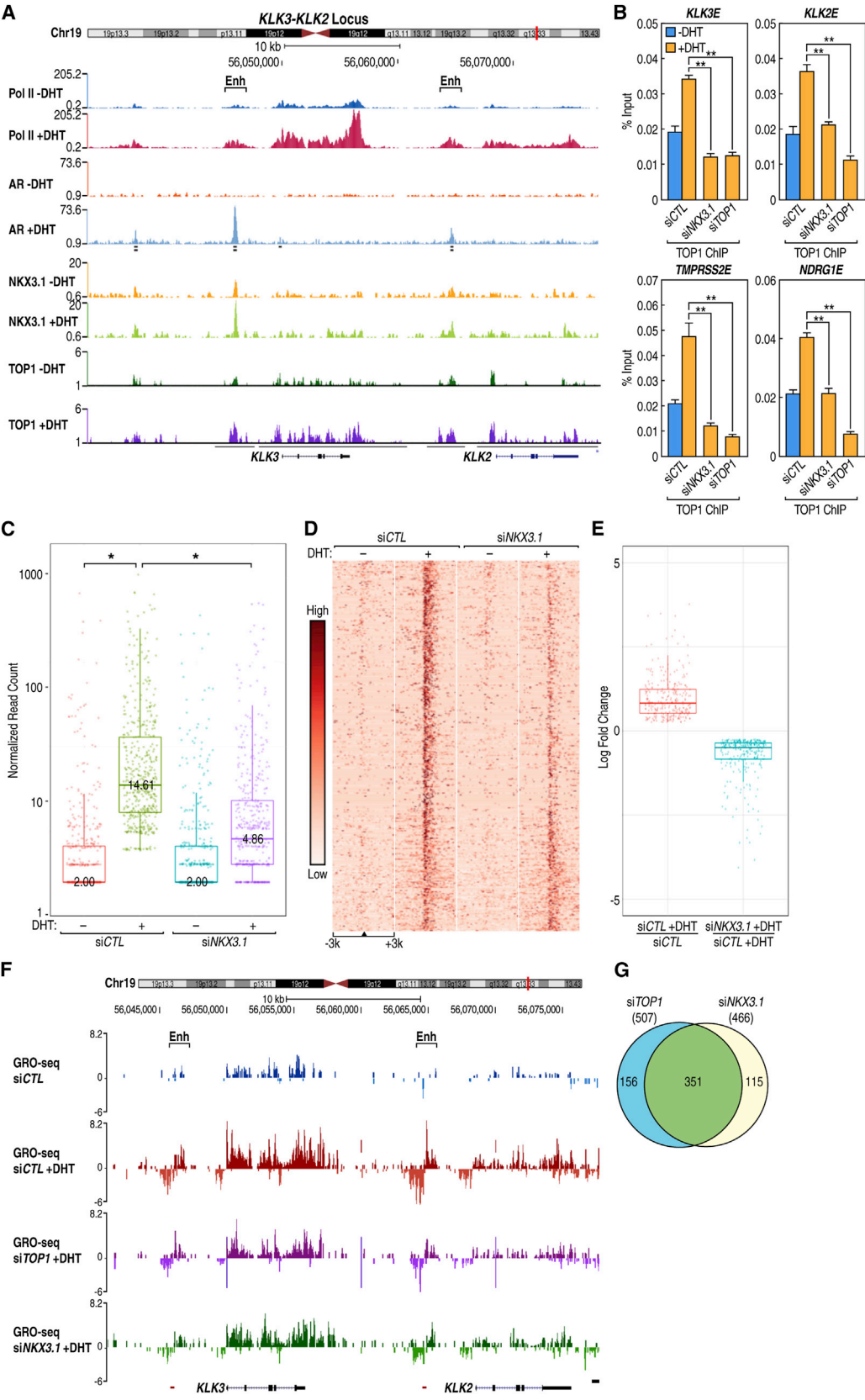
(E) Box plot: siTOP1 reduced transcription at ~80% of DHT-upregulated AR enhancers. * $p \leq 2.2 \times 10^{-16}$ (Wilcoxon test).

(F) Box plot: the response to DHT of 368 DHT-upregulated genes was reduced after *TOP1* knockdown by siRNA.

(G) Knockdown of *TOP1* affects the induction of both eRNA and mRNA. LNCaP cells, hormone-starved for 1 day and transfected with the indicated siRNA, were stimulated with 100 nM DHT for 1 hr (eRNA) or 5 hr (mRNA) 48 hr posttransfection. qRT-PCR was performed with SYBR Green using reverse-transcribed RNA. Data represent mean \pm SD ($n = 3$), ** $p < 0.01$.

(H) Recruitment of ATR to the *KLK3* and *KLK2* enhancers following DHT stimulation of starved cells. Data represent mean \pm SD ($n = 3$). ** $p < 0.01$.

See also Figure S1 and Table S1.



(legend on next page)

enhancers with significantly upregulated eRNAs (Figure 1C), which is the best mark of activated enhancers (Hsieh et al., 2014; Li et al., 2013a), among which 477 (~74%) were noted to have increased TOP1 occupancy in response to ligand at 30 min (Figure 1D), and virtually all appear to exhibit DHT-increased TOP1 binding at 15 min (Figure S5C). Because TOP1 has been shown to affect the transcriptional activity of RNA Pol II (Kretschmar et al., 1993), we decided to investigate whether knockdown of TOP1 would alter eRNA synthesis from the androgen-regulated enhancers. Knockdown of endogenous TOP1 by small interfering RNA (siRNA) revealed that eRNA induction was reduced in at least 79% (507 of 644) of AR-regulated enhancers (Figures 1C and 1E), accompanied by a decrease in the induction of 368 coding target genes in the experiment shown (Figure 1F), with similar results in repeat experiments. Ninety-two percent of DHT-induced eRNAs were upregulated more than 2-fold, with fold change average of 7.6 times. Analysis of 100 randomly selected housekeeping genes not regulated by DHT in our GRO-seq experiments confirmed that the specific siRNAs used for this study had no effect on their expression (Table S1).

To validate all major mechanistic points in this study, we chose four enhancers gene pairs. Three of these enhancers (KLK3E, KLK2E, and TMPRSS2E) are validated by previous studies (Andreu-Vieyra et al., 2011; Clinckemalie et al., 2013; Hsieh et al., 2014). The fourth one, NDRG1E, meets the criteria of others. It is an AR-bound element located not too far away from the *NDRG1* gene TSS (−29 kb), it is H3K4me1⁺, H3K27Ac⁺, and following hormone stimulation the transcription unit produces DHT-dependent, bidirectional eRNA, making it a strong candidate. Using these enhancer sites, we found that recruitment of the nuclear receptor coactivators (p300 and SRC-1) at AR enhancers was diminished after siTOP1 (Figure S1E). Thus, TOP1 knockdown attenuated the induction of eRNA (1 hr DHT treatment) and the production of mRNA of the corresponding target genes 5 hr after ligand addition (Figures 1G and S1F). Importantly, the fold induction (with or without DHT) was similar between independent experiments in which eRNA levels were measured. Surprisingly, we noted that ATR (Ataxia telangiectasia and Rad3-related), a protein involved in DNA damage repair, was recruited to AR-regulated enhancers at ~15 min following addition of ligand (Figures 1H and S1G). Together, these data identify TOP1-bound genomic regions that bear enhancer marks and produce eRNA in a DHT-dependent manner. Knockdown of TOP1 reduces production of eRNA and coding gene RNA for most of these AR-regulated target genes.

NKX3.1 and TOP1 Co-Occupy Enhancer Binding Sites and Regulate the AR Transcription Program

NKX3.1 is an androgen-regulated transcription factor (Bhatia-Gaur et al., 1999), which is a highly selective and specific marker of metastatic prostatic adenocarcinoma (Gurel et al., 2010). NKX3.1 has been found to interact with TOP1 to enhance formation of the TOP1-DNA complex and increase TOP1 nicking of DNA (Bowen et al., 2007). In fact, TOP1 activity in prostates of *Nkx3.1*^{+/-} and *Nkx3.1*^{-/-} mice is reduced compared with wild-type mice, but not in other organs that do not express *Nkx3.1* (Bowen et al., 2007). Overlap of the reported NKX3.1 ChIP-seq data set (Tan et al., 2012) with that of AR and TOP1 revealed that NKX3.1 occupancy was highest at AR enhancers, with NKX3.1 binding sites located over regions with increased TOP1 binding (Figures 2A and S2A). We observed that AR and TOP1 started to be recruited to AR enhancers within a few minutes after DHT stimulation. Interestingly, siRNA-mediated knockdown of cellular NKX3.1 inhibited recruitment of TOP1 at enhancers of DHT-regulated genes at 5 min following DHT stimulation (Figure 2B), in line with the previous data suggesting that NKX3.1 is needed for the formation of the TOP1-DNA cleavage complex (Bowen et al., 2007). We observed that, following NKX3.1 knockdown in LNCaP cells, the DHT-dependent upregulation of ~70% enhancer eRNAs was significantly reduced (Figures 2C, 2D and S2C). We also noted significant reduction in the expression levels of 273 DHT-upregulated genes (Figure 2E), exemplified for two representative genes (Figure 2F). Additionally, knockdown of TOP1 and NKX3.1 reduced DHT upregulation of eRNA at the same 351 AR enhancers in these experiments (Figure 2G), apparently without affecting AR recruitment to the enhancer-binding sites (Figure S2E). Together, these experiments demonstrate that NKX3.1 and TOP1 binding occurs at a subset of DHT-regulated enhancers, and the knockdown of either diminishes transcription in response to ligand.

Catalytic Activity of TOP1 Is Required for DNA Nicking and Enhancer Activation

Based on its mechanism of action as a DNA nickase, by which TOP1 forms a covalent intermediate with DNA and possesses intrinsic DNA ligase activity (Pommier et al., 1998; Champoux, 2001), it would be difficult to detect any such transient nick by available methods. Indeed, despite extensive attempts to detect such a nick in enhancers by primer extension approaches, only a few examples could be clearly visualized. Thus, using this approach to investigate whether AR-regulated enhancers might be the sites of DNA scission by the activated TOP1, we chose the

Figure 2. NKX3.1 and TOP1 Co-Occupy a Subset of AR Enhancers and Co-Regulate the Enhancer Program

- (A) The UCSC genome browser screenshot displaying a direct overlap between AR, NKX3.1, and TOP1 binding at enhancers of *KLK3* and *KLK2* genes. Regions with increased (after DHT) TOP1-binding (except regions present in the background control) are underlined.
- (B) Knockdown of *NKX3.1* prevents TOP1 from binding at AR-regulated enhancers; siCTL-, siNKX3.1-, and siTOP1-treated cells were stimulated with DHT for 5 min. Chromatin immunoprecipitation was performed with an antibody against TOP1. Data represent mean \pm SD (n = 3). **p < 0.01.
- (C) Knockdown of *NKX3.1* by siRNA affects the induced transcription of ~69% of the regulated eRNAs. *p $\leq 2.2 \times 10^{-16}$ (Wilcoxon test).
- (D) Heatmap of AR enhancers sorted from most-to-least affected by siNKX3.1.
- (E) siNKX3.1 reduces induced transcription of 273 genes in this experiment determined by GRO-seq.
- (F) The UCSC genome browser screenshot showing the *KLK3-KLK2* locus. Knockdown of TOP1 or NKX3.1 by siRNA reduces eRNA and genic RNA induction.
- (G) Knockdown of either TOP1 or NKX3.1 affects induction of the same 351 eRNAs in the same experiment, as measured by GRO-seq.
- See also Figure S2 and Table S1.

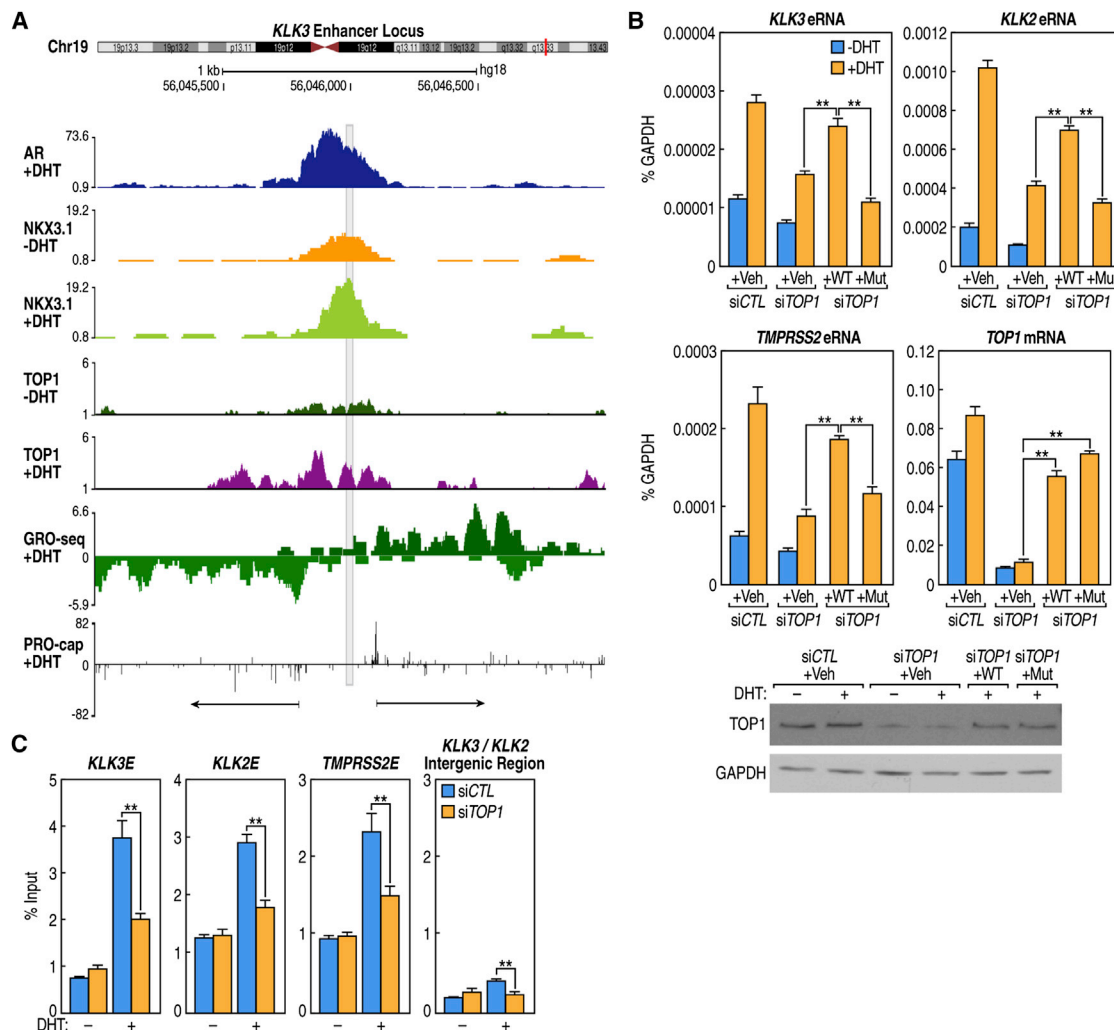


Figure 3. TOP1 Recruits to AR-Regulated Enhancers and Nicks the DNA

(A) UCSC browser screenshot displaying the *KLK3* enhancer. Arrows indicate the PRO-caps representing (putative) eRNA TSS that flank the NKX3.1 peak. (B) eRNA readout assay showing that Tyr723 of TOP1 is required for eRNA induction. LNCaP cells were hormone-starved for 24 hr, transfected with siRNA to knock down *TOP1*, and then electroporated with empty expression vector (Veh), wild-type TOP1 (WT), or the Y723F-TOP1 mutant (Mut) before treatment with either ethanol or DHT for 1 hr. eRNA for *KLK2*, *KLK3*, and *TMPRSS2* gene enhancers was quantified by RT-PCR. TOP1 mRNA and protein levels are also shown. qPCR data show mean \pm SD ($n = 3$). ** $p < 0.01$.

(C) Knockdown of endogenous *TOP1* affects nick/break formation as measured by incorporation of biotin 11-dUTP at selected AR enhancers after 10 min DHT treatment. Data represent mean \pm SD ($n = 3$). ** $p < 0.01$. See also Figure S3 and Table S2.

KLK3 enhancer as a model. We examined a region overlapped by the AR and NKX3.1 peaks and flanked by two “precision nuclear run-on transcription initiation sites” (PRO-caps), which mark the transcription initiation sites at high resolution (Kwak et al., 2013), noting that PRO-cap sites could be located on AR-regulated enhancers following hormone stimulation (Figure 3A, Table S2). Primer extension analysis of both DNA strands with [γ^{32} P]-ATP-labeled oligonucleotides yielded several termination products consistent with a series of closely spaced DNA nicks; the strongest band that became accentuated in response to DHT was seen on the lower strand, in support of the notion that it may be one of the major TOP1 binding/scission

sites (Figure S3B). Moreover, detailed PRO-cap analysis to locate the precise start sites revealed that the RNA cap sites located on average ~ 134 bp away from the center of the AR peak (Figure S3C) are occupied by TOP1 (Figure S3D) and, as shown in GRO-seq experiments, these transcripts continue to the end of eRNA-encoding sequence; however, for the majority (75th percentile) of these transcripts, the GRO-seq signal starts to fade away after 1,000 bp from the TSS/cap site.

As another approach to infer the possibility of TOP1 DNA nickase actions in activation of AR-dependent enhancer, we sought to mutate TOP1. TOP1 enzymatic activity depends on Tyr723 to relax superhelical DNA (Madden and Champoux, 1992).

Specifically, Tyr723 of TOP1 initiates the nucleophilic attack on the backbone scissile phosphate resulting in nicked DNA and a phosphodiester link between the tyrosine and 3' phosphate (Champoux, 2001; Pommier et al., 2010). Subsequently, the covalent intermediate is religated with concomitant release of Tyr723 from the DNA (Champoux, 2001; Stewart et al., 1998). We therefore tested whether the Y723F TOP1 mutant could rescue the defect caused by *TOP1* knockdown. For this purpose, endogenous *TOP1* was knocked down with specific siRNA, and either the wild-type or the Y723F TOP1 mutant was then expressed in LNCaP cells. Analysis of the enhancer RNA after 1 hr DHT treatment revealed that the wild-type TOP1 largely reinstated eRNA induction, whereas the catalytically inactive mutant failed to do so (Figure 3B). The incomplete rescue with the wild-type construct most likely reflected the fact that not all cells could be efficiently electroporated with the DNA expression vectors, probably because LNCaPs are notoriously difficult to transfect with conventional cationic liposome reagents. Interestingly, wild-type TOP1 relaxes supercoiled DNA only in the presence of NKX3.1, whereas the active site mutant does not at all (Bowen et al., 2007), consistent with the presence of TOP1 on AR-bound enhancers. These findings are of particular interest based on previous in vitro transcription system analyses. TOP1 has been shown to be essential for transcriptional activation in a system containing RNA polymerase II and other cofactors (Kretschmar et al., 1993; Merino et al., 1993; Shykind et al., 1997), but, in these artificial in vitro transcription systems, the Y723F mutant did not block the transcriptional activity of the complex at promoters. Therefore, in this context, TOP1 was proposed to modulate transcription by changing the conformation of DNA at the promoter or via interactions with TBP/TFIID (Kretschmar et al., 1993; Merino et al., 1993; Shykind et al., 1997). In contrast, on AR-regulated enhancers, the nicking activity of TOP1 appears to be required for its effects on eRNA transcription.

Incorporation of labeled nucleotide by terminal deoxynucleotidyl transferase (TdT) has been considered to label both DNA nicks and double-stranded DNA (dsDNA) breaks (Gavrieli et al., 1992); hence, we also employed this assay on specific enhancer sites to assess incorporation of biotin 11-dUTP in response to DHT. Therefore, we fixed the cells with *Streck Cell Preservative* (Ju et al., 2006), a formulation shown not to cause DNA breaks during the fixation process. Biotin 11-dUTP incorporation with TdT was observed at 10 min following addition of DHT hormone at the several enhancers tested, and this was strikingly reduced after *TOP1* knockdown (Figure 3C). Together, these data suggested that TOP1 recruitment to enhancers co-occupied by AR and NKX3.1 occurred at regions proximal to transcription initiation sites and caused single-stranded DNA (ssDNA) nicks, although the possibility of a dsDNA break cannot be ruled out, especially as an unligated nick can be converted to a DSB for subsequent processing by the DSB repair pathway (Davis and Maizels, 2014).

Involvement of MRE11 in the Regulation of the AR Program

The MRN complex, composed of the meiotic recombination 11 (MRE11), RAD50, and Nijmegen breakage syndrome 1 (NBS1), is central to the DNA damage response (DDR) pathway that is

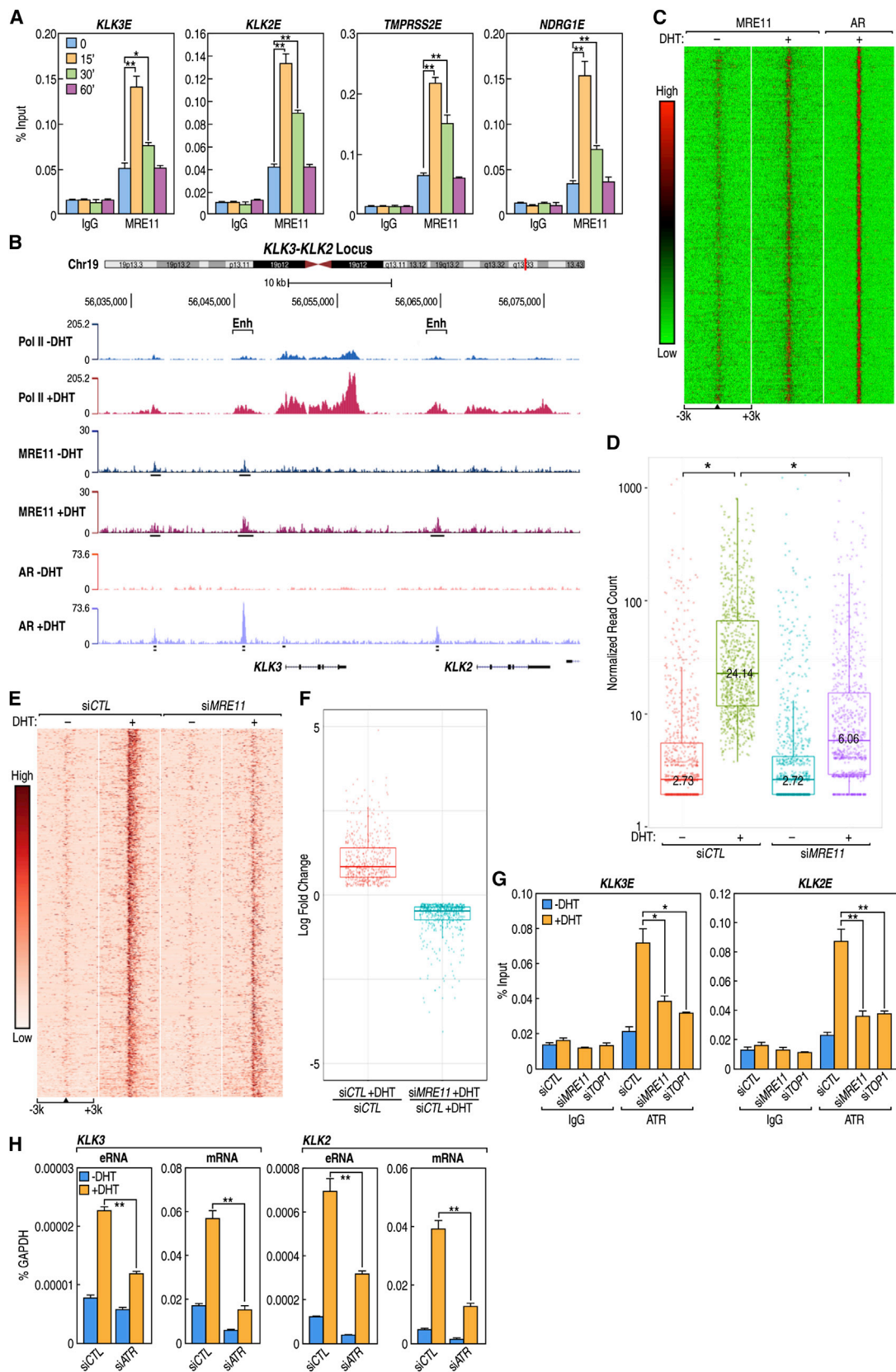
initiated upon recognition of the DNA breaks by sensor proteins (Stracker and Petrini, 2011). MRE11 regulates DNA repair by recruitment of DNA-repair proteins that load onto the chromatin at the site of the break (Price and D'Andrea, 2013).

Recent evidence shows that cleavage of the covalent 3' phosphotyrosyl-DNA bonds that join TOP1 to the DNA backbone by MRE11 generates a product carrying a 3'-phosphate end that MRE11-RAD50 can resect in an ATP-regulated reaction, producing a 3'-hydroxyl that can prime repair synthesis (Hamilton and Maizels, 2010; Sacho and Maizels, 2011). Interestingly, the p300 transcriptional coactivator physically interacts with all three members of the MRN complex (Jung et al., 2005).

Based on these considerations and the results in Figure 3, we investigated whether MRE11 was present at AR-regulated enhancers. Kinetic ChIP experiments using a specific antibody (Figures 4A and S4A) revealed that MRE11 recruitment at enhancer-binding sites peaked at 15 min of DHT treatment. On performing ChIP-seq, we identified 19,886 loci in the (–) hormone control and 30,636 loci in the cells treated with DHT for 15 min, observing that MRE11 sequencing tag density at enhancers increased with DHT treatment (Figures 4B, 4C, and S4B). We also observed similar recruitment of the RAD50 component of the MRN complex (Figure S4C). Genome-wide analysis showed indistinguishable alterations in the number of tags over promoters of these genes in response to DHT treatment, although a small increase in MRE11 occupancy at promoters of select DHT-regulated genes (e.g., *KLK3*, *KLK2*, *NDRG1*, and *TMPPSS2*) could be detected by ChIP-qPCR after DHT treatment (data not shown). GRO-seq analysis of nascent transcription revealed that induction of ~89% of detectable enhancer eRNAs induced by DHT were inhibited by *MRE11* knockdown (Figures 4D and 4E). In addition, expression of 510 induced coding genes was reduced (Figure 4F). Knockdown of *RAD50* caused a similar effect on eRNA and mRNA expression levels (Figure S4D). Given the role of ATR in sensing single-strand DNA breaks, we also investigated the potential functional role of ATR following DHT. We found that ATR is rapidly recruited, by 15 min, to AR-bound enhancers after DHT (Figure 1H). This is of functional significance, because knockdown of either *MRE11* or *TOP1* caused dramatic decrease in ATR recruitment to enhancers (Figure 4G) and a reduction of DHT-induced enhancer and gene transcription (Figure 4H).

Recruitment of Components of DDR to AR-Regulated Enhancers

Indeed, a mechanism that could be involved in the repair of single-strand nick would be the base excision repair pathway (BER), to process nicks that evaded TOP1 ligase activity. Therefore, we investigated whether factors involved in this or other DNA damage repair pathways might also be recruited to AR-regulated enhancers. We performed kinetic ChIP experiments using antibodies against phospho-ATM (Ataxia telangiectasia mutated), Ku80 (part of the Ku heterodimer that binds to double-strand DNA break ends), Exonuclease 1 (EXO1), the Bloom syndrome DNA helicase (BLM), and DNA ligase IV (LIGIV). Additionally, we used antibodies to proteins involved in the base excision repair pathway, including XRCC1 (X-ray repair cross-complementing protein 1), DNA polymerases β and ϵ ,



(legend on next page)

and Ligase I, observing an orderly and reproducible kinetics of recruitment after hormone treatment at enhancers including *KLK3*, *KLK2*, *TMPPRSS2*, and *NDRG1* (Figures 5 and S5A), as well on other DHT-upregulated enhancers identified by the GRO-seq (Figure S5B). Although TOP1 and ATR were essentially recruited simultaneously at enhancers at 15 min, XRCC1 was recruited between 15 and 30 min, consistent with the recruitment of base excision repair pathway machinery that could process any unligated nicks. Interestingly, DNA ligase IV showed maximum occupancy after 30 min, whereas pATM (p-S1983), Ku80, EXO1, BLM, and DNA ligase I were maximally recruited to enhancers ~60 min post DHT treatment (Figures 5 and S5A), indicating recruitment of multiple DNA-repair factors that have been conventionally considered to function in DNA damage repair (Nimonkar et al., 2011). The sequence of events would be consistent with resolving any unligated DHT/TOP1-induced ssDNA nicks; the DDR machinery primarily recruited as a “safety net” against any DNA breaks that are not sealed by TOP1. From our data, the machineries of transcription and DNA damage repair seem to be intrinsically linked.

DISCUSSION

Regulated gene expression has been a subject of intense investigation over the past few decades, yet the precise mechanisms by which enhancers orchestrate tissue-specific programs with such an astonishing precision remain unclear. In particular, the finding that enhancers are also regulated transcription units, encoding eRNAs, has added to the mystery and raised new questions about how the subsequent topological strain on enhancers is handled. Both eRNA synthesis and nucleosome depletion at enhancers are potential sources of topological strain. Advancing RNA polymerase can generate both positive and negative supercoils. The amount of supercoiling is potentially enormous given that a positive and a negative supercoil is generated for every 10 bp transcribed and that the length of an eRNA transcript is typically 1–2 kbp in length. Indeed, it has been estimated that approximately seven supercoils may be generated by the transcribing polymerase per second, and that these supercoils can propagate >1 kbp from the transcription start site (Kouzin et al., 2013). At the same time, the depletion of histones from enhancers releases unconstrained negative supercoils, which, in principle, can parse to a change in DNA twist or unwinding to facilitate transcription and/or to a change in writhe that impedes transcription factor binding. To relieve torsional stress, it is tempting to predict that cells might employ

actions of DNA topoisomerases, including topoisomerase I as an integral component of regulated enhancer transcription.

Here, we have elucidated the operation of just such a mechanism in prostate cell-specific enhancer activation by androgen receptor, using the LNCaP cancer cell line as a model. In a sense analogous to the role of TOP1 at origins of replication (Simmons et al., 1998; Tsao et al., 1993), we show here that this DNA nickase is rapidly recruited to a large cohort of AR/NKX3.1-occupied enhancers to putatively activate the enhancers and relieve torsional stress due to ongoing transcription (Figure 6). Our results are consistent with observations that, in yeast cells, Top1/Top2 play a role in the activation of genes characterized by high transcriptional plasticity (Pedersen et al., 2012). However, the beneficial effects of TOP1 have to be weighed against the negative effects of retention of TOP1 as an obstacle to further transcription and the deleterious effects of a single-strand nick if it is not quickly sealed by TOP1 itself, or repaired by the base excision pathway. Unrepaired nicks could lead to the formation of DNA double-strand breaks (DSB) as, for example, when a replication fork runs into and collapses at a nick (Kuzminov, 2001; Wimberly et al., 2013). It has also been suggested that a codirectional collision between the replisome and backtracked RNA polymerase transcription elongation complexes leads to DNA double-strand breaks (Dutta et al., 2011). Thus, one important role for the MRN complex and other components of the DDR machinery that we observe recruited to the TOP1-bound enhancers might be for the removal of any “stalled” TOP1 from the DNA substrate, as well as repair of any possible DNA breaks that might occur despite TOP1 or the BER actions (Hamilton and Maizels, 2010; Sacho and Maizels, 2011; Davis and Maizels, 2014).

TOP1 activity is likely to be modulated by factors other than NKX3.1, suggesting that the mechanism we describe here may not be restricted to prostate cells. In this regard, it has been shown that the catalytic activity of TOP1 is stimulated by large T antigen during unwinding of the SV40 origin (Simmons et al., 1998) and overexpression of the antigen rendered LNCaP cells androgen-independent for cell-cycle progression (Knudsen et al., 1998). This raises the possibility that activation of TOP1 catalytic activity may, in part, trigger a switch to androgen independence. The Werner syndrome helicase, WRN, has also been found to enhance the ability of TOP1 to relax negatively supercoiled DNA and specifically stimulate the religation step of the relaxation reaction (Laine et al., 2003). It is therefore not unlikely that there exist other, yet undiscovered, activators of TOP1 catalytic activity to regulate eRNA synthesis and gene expression

Figure 4. MRE11 Regulates the AR Transcription Program

- (A) Recruitment of MRE11 to the selected DHT-regulated AR enhancers. Data points show mean \pm SD (n = 3). *p < 0.05, **p < 0.01.
 (B) MRE11 binding (sequencing tags density) increases over AR enhancers in a DHT-dependent manner (*KLK3* and *KLK2* genes shown).
 (C) Distribution of MRE11 and AR binding (sequencing tag density) centered over AR-enhancer binding sites with DHT-induced eRNA.
 (D) *MRE11* knockdown reduces eRNA expression levels of 89% of DHT-upregulated eRNAs. *p $\leq 2.2 \times 10^{-16}$ (Wilcoxon test).
 (E) Heatmap for AR enhancers sorted from the most downregulated by siMRE11 at the top to the least, at the bottom.
 (F) Boxplot showing 510 genes, where DHT-induced upregulation of transcription (determined by GRO-seq) was reduced by *MRE11* knockdown.
 (G) Knockdown of either *MRE11* or *TOP1* affects recruitment of ATR at enhancers following hormone stimulation of the starved cells, measured after 15 min DHT stimulation. Data show mean \pm SD (n = 3). *p < 0.05, **p < 0.01.
 (H) siATR affects induction of eRNA (1 hr DHT) and mRNA (5 hr DHT treatment) of the corresponding gene. Data are the mean \pm SD (n = 3). *p < 0.05, **p < 0.01. See also Figure S4 and Table S1.

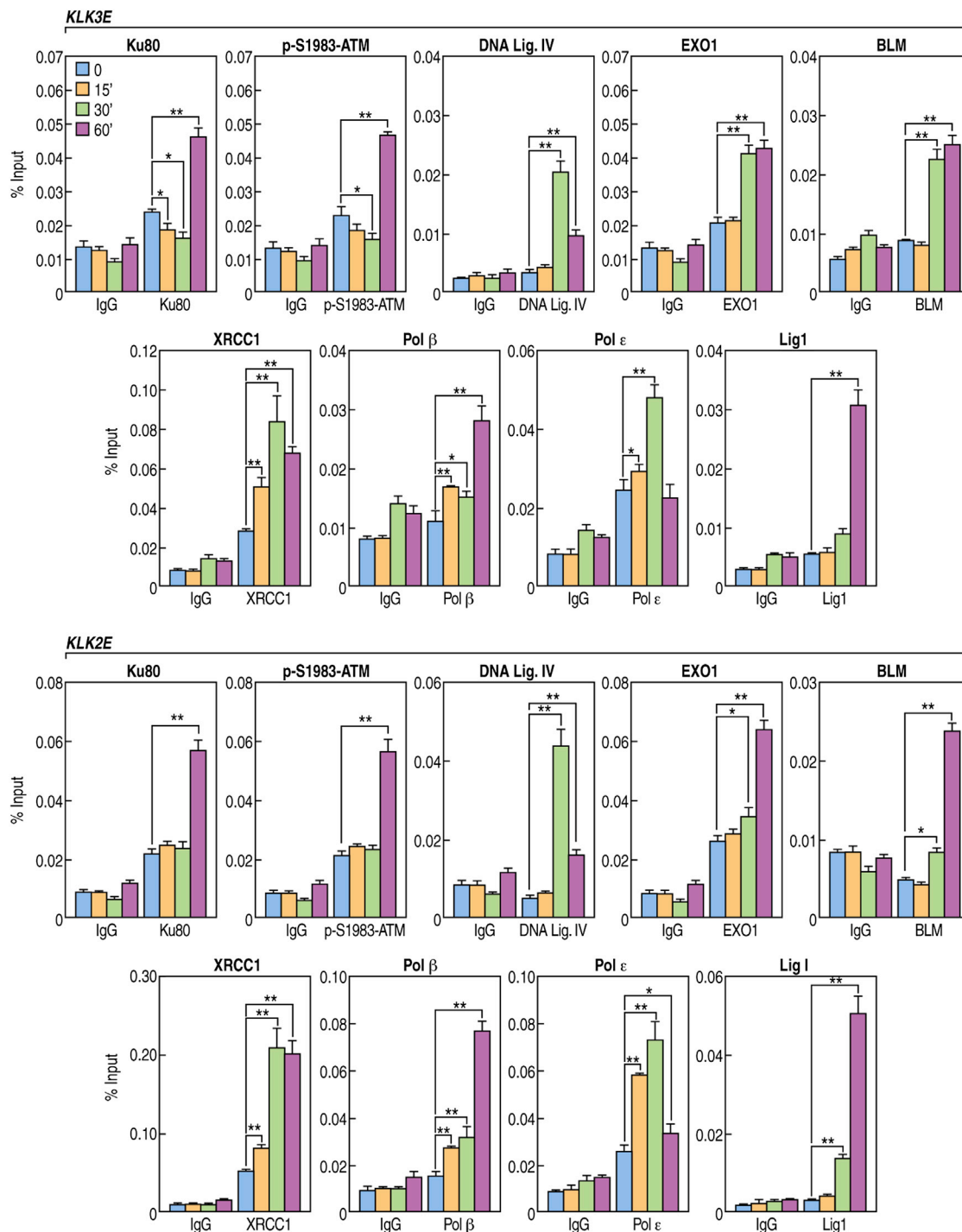


Figure 5. Canonical DNA Damage/Repair Machinery Components Recruit to AR-Regulated Enhancers

Kinetic recruitment of factors implicated in the DNA damage response (DDR) to AR enhancers. All kinetic ChIP experiments were performed at least twice with cells of similar passage number to ensure data reproducibility. Data shown as mean \pm SD ($n = 3$). * $p < 0.05$, ** $p < 0.01$. See also Figure S5.

programs. Alternatively, there may be other DNA nickases that initiate enhancer activation in tissues other than prostate, in signal-dependent manner, and that the activities of those nickases are modulated by enhancer-bound factors.

Although the finding that ligand-dependent enhancer activation strategy would involve a DNA nick may seem counterintuitive in terms of cellular integrity, it is noteworthy that cellular integrity is threatened daily by endogenous and extracellular

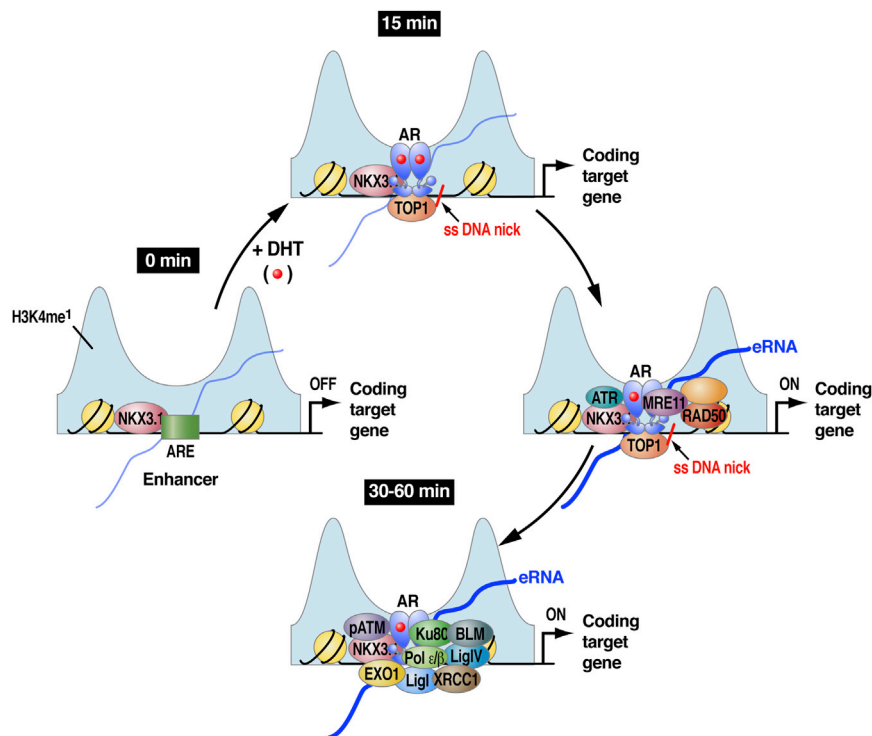


Figure 6. A Model for TOP1-Mediated Activation of the AR Enhancer

Following androgen stimulation, AR and DNA topoisomerase I recruit to the enhancer region, premarked by the NKX3.1 pioneer transcription factor. NKX3.1 to TOP1 stimulates enzymatic activity of the topoisomerase, resulting in nicking of DNA on a single strand, followed by recruitment of ATR, XRCC1, and the MRN complex components (MRE11/RAD50). After dismissal of TOP1, ATR, and the MRN, additional components of DNA-repair machinery recruit to the activated enhancer. The thin blue line indicates the presence of low levels of residual eRNA, not totally eliminated by hormone starvation, whereas the thick blue line represents induced bidirectional eRNA produced by the transcription unit.

the DNA-dependent protein kinase catalytic subunit and was highly selective. In addition, there was no detectable association with the initiating isoform of RNAPII or with the general transcription initiation factors. The authors concluded that association of Ku80 with transcription sites is important for maintenance of global transcription levels, because functional

agents that lead to the formation of single- and double-strand DNA breaks. For instance, the estimated number of single-strand breaks and spontaneous base losses in nuclear DNA together with other types of spontaneous damage may reach 10^5 lesions per cell per day (Hoeijmakers, 2009), yet the cells are programmed to survive. To maintain genomic integrity, cells constantly engage the DNA-repair machinery. As such, the usage of a programmed DNA nicking/repair strategy in regulated transcription to relieve torsional stress and activate transcription in this case, while apparently surprising, is in keeping with growing evidence that components of DNA damage machinery do participate in transcriptional regulation. For instance, Reinberg and colleagues demonstrated that human RNA polymerase II complex contains components with roles in DNA repair, including Ku70, Ku80, and DNA Pol ϵ (Maldonado et al., 1996), and Kung and colleagues (Mayeur et al., 2005) have identified heterotrimeric DNA-dependent protein kinase subunits: Ku70, Ku80, and DNA-PKcs, as well as poly(ADP-ribose) polymerase as proteins associated with the C-terminal domain of AR and demonstrated that, in LNCaP cells, Ku70 and Ku80, recruited to the *KLK3* promoter and enhancer in a hormone-dependent manner. Interestingly, Ku70 and Ku80 can function outside of the Ku heterodimer that loads on double-strand DNA breaks. Hasty and colleagues have shown that Ku80 deletion impairs the base excision pathway (BER) at the initial lesion recognition/strand scission step, arguing that free Ku70 and free Ku80, but not the Ku heterodimers, associate with apurinic/apyrimidinic (AP) sites that BER corrects (Li et al., 2013b; Choi et al., 2014). Moreover, Mo and Dynan showed that, in normally growing human cells, Ku80 associated with RNA polymerase II elongation sites. This association occurred independently of

disruption of a discrete C-terminal domain in the Ku80 subunit inhibited transcription in vitro and in vivo (Mo and Dynan, 2002). Importantly, LigIV, like Ku80, is commonly associated with the NHEJ pathway, but its active site has been found to be highly permissive and capable of ligating atypical DNA substrates, including nicks with gaps (Gu et al., 2007). Interestingly, in the absence of RNase H2, the suppression of mutations arising from misinsertion of ribonucleoside monophosphates (rNMP) during DNA replication involves Top1-mediated cleavage at an rNMP, followed by unwinding of DNA by Srs2 and digestion by Exo1 (Potenski et al., 2014). Also, earlier studies showed that TOP1 enhanced TFIIID-TFIIA complex assembly during activation of transcription; however, in these biochemical studies, the catalytic activity of TOP1 was not essential to activate transcription from promoters. It is also interesting to note that the AR itself has been shown to transcriptionally regulate a network of DNA-repair genes, including those implicated in DNA damage sensing (MRE11, NBN, and ATR), nonhomologous end joining (XRCC4 and XRCC5), homologous recombination (RAD54B and RAD51C), mismatch repair (MSH2 and MSH6), base excision repair (PARP1 and LIG3), and the Fanconi pathway (FANCC, FANCI, and USP1) (Polkinghorn et al., 2013). Moreover, p53 itself binds enhancers and regulates eRNA synthesis for transcription enhancement of neighboring genes (Melo et al., 2013).

Together, the recruitment of DNA damage response machinery in specific transcriptional regulatory events is an emerging theme, from the regulation of pluripotency in embryonic stem cells by the trimeric XPC-nucleotide excision repair complex (Fong et al., 2011) to the regulation of human RAR β 2 gene via XPG induced DNA breaks at the promoter region (Le May et al., 2012). Moreover, experiments with yeast have revealed

that the Rad1^{XPF}/Rad10^{ERCC1}, Mms4^{Emi1} orthologs can catalyze the endonucleolytic cleavage of DNA immediately upstream from the Top1-DNA adduct (Pommier et al., 2010). Indeed, permissive chromatin architecture seems to be a crucial requirement for transcription initiation events (Fong et al., 2013). Although these events are quite distinct from the TOP1-dependent regulatory events described in the present manuscript, they do suggest a common usage of the DNA damage repair machinery to regulate gene transcription.

EXPERIMENTAL PROCEDURES

Cell Culture

LNCAp cells were purchased from ATCC and maintained in RPMI-1640 medium (Life Technologies) supplemented with 10% fetal bovine serum (Omega Scientific), 2 mM L-glutamine, and penicillin/streptomycin. For kinetic ChIP experiments, cells were starved in phenol-free DMEM (Lonza) supplemented with 5% charcoal:dextran stripped fetal bovine serum (Omega Scientific) for 72 hr. Cells were synchronized with 2.5 μ M α -amanitin (Sigma) for 2 hr, washed twice with PBS, and released. A total of 100 nM 5 α -Dihydrotestosterone (DHT, Sigma) was added to the starvation media to stimulate the cells.

Small Interfering RNA

siRNA-mediated knockdown was achieved by transfecting cells with Lipofectamine 2000 and specific siRNAs. The following siRNAs were used for this study: AllStars Neg. Control siRNA (1027281) was from QIAGEN. Human ON-TARGETplus SMARTpool siRNAs against TOP1 (L-005278-00-0020), MRE11 (L-009271-00-0020), and RAD50 (L-005232-00-0005) were purchased from Dharmacon. Single interfering RNAs targeting AR (SASI_Hs01_00224483, SASI_Hs01_00224484), TOP1 (SASI_Hs02_00335354, SASI_Hs01_00047440), ATR (SASI_Hs01_00176270, SASI_Hs01_00176271), and NKX3.1 (SASI_Hs02_00341026, SASI_Hs01_00018365) were obtained from Sigma. Multiple siRNAs were used during the course of the study to confirm data reproducibility.

For transfection, LNCAp cells were seeded on dishes in RPMI-1640 supplemented with 10% FBS and allowed to attach overnight. The following day, the cells were washed twice with PBS and fed with phenol-free DMEM supplemented with 5% charcoal:dextran FBS. One day later, the cells were transfected using Lipofectamine 2000 and 20 pmol ml⁻¹ siRNA diluted in Opti MEM reduced serum media without phenol red (Life Technologies). The transfection media was removed after 16 hr incubation, and the cells were washed twice with PBS. Fresh, phenol-free DMEM supplemented with 5% charcoal:dextran FBS and penicillin/streptomycin was added to the dishes. Cells were harvested 48–72 hr posttransfection. All siRNAs used in this study were validated by vendors or by us and used only if providing >70% knockdown efficiency. Relative quantities of gene expression level were normalized to the *GAPDH* gene. The relative quantities of ChIP samples were normalized by individual inputs, respectively.

ChIP-qPCR

Chromatin immunoprecipitation experiments were done as previously described (Garcia-Bassets et al., 2007). All ChIPs and qPCRs were repeated at least thrice and representative results were shown. p values were calculated by using a two-tailed Student's t test.

GRO-Seq and PRO-Cap

Global run-on sequencing (GRO-seq) was performed as detailed (Wang et al., 2011), and precision nuclear run-on sequencing of transcription initiation sites (PRO-cap) was performed as described (Kwak et al., 2013).

Antibodies

AR (N-20), TOP1 (H-300), ATR (N-19), RAD50 (H-300), XRCC1 (H-300), BLM (H-300), DNA Ligase I (C-21), DNA Ligase IV (H-300), DNA POL β (C-21), DNA POL ϵ 3/CHRA17 (N-15), p300 (C-20), SRC-1 (M-341) were from Santa Cruz Biotechnology. MRE11 (Ab397) and p-S1983-ATM (Ab2888)

were obtained from Abcam. Ku80 (A302-627A) and EXO1 (A302-639A) were purchased from Bethyl Laboratories. See also [Extended Experimental Procedures](#).

ACCESSION NUMBERS

The Gene Expression Omnibus accession number for all sequencing data reported in this paper is GSE63202.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, five figures, and two tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.12.023>.

AUTHOR CONTRIBUTIONS

J.P. and M.G.R. conceived the project. J.P., A.K.A., and M.G.R. designed experiments. J.P. performed most of the experiments reported. P.K. performed bioinformatics analyses. W.L. performed GRO-seq experiments, and Y.T. performed the modified PRO-cap experiments. T.S. carried out screening experiments, while Z.L. generated biotin-TOP1 stable cell line and carried out TOP1 ChIP-seq. J.Z. and K.A.O. assisted in deep-sequencing library preparations and sequencing. J.P., A.K.A., and M.G.R. wrote the manuscript with input from P.K.

ACKNOWLEDGMENTS

We thank Dr. Bogdan Tanasa and Daria Merkurjev for help with statistical analyses, Charles Nelson for cell-culture assistance, Tara Rambaldo for help with cell-cycle analysis, and Janet Hightower for assistance with figure preparation. We are particularly grateful to Dr. E. Peter Geiduschek for a critical reading of the manuscript and Drs. Kalotina Machini and Patricia Cortes for discussions and helpful suggestions and thank the members of the M.G.R. laboratory for their comments on the work. This work was supported by grants from NIH and NCI to M.G.R. (DK018477, HL065445, NS034934, GM104459, DK039949, DK074868, DK097748, and CA173903). J.P. was supported, in part, by a NIDDK Mentored Research Scientist Development Award (K01DK080180). M.G.R. is an Investigator with the HHMI.

Received: July 23, 2014

Revised: October 28, 2014

Accepted: December 10, 2014

Published: January 22, 2015

REFERENCES

- Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al.; FANTOM Consortium (2014). An atlas of active enhancers across human cell types and tissues. *Nature* 507, 455–461.
- Andreu-Vieyra, C., Lai, J., Berman, B.P., Frenkel, B., Jia, L., Jones, P.A., and Coetzee, G.A. (2011). Dynamic nucleosome-depleted regions at androgen receptor enhancers in the absence of ligand in prostate cancer cells. *Mol. Cell. Biol.* 31, 4648–4662.
- Bhatia-Gaur, R., Donjacour, A.A., Scialvolino, P.J., Kim, M., Desai, N., Young, P., Norton, C.R., Gridley, T., Cardiff, R.D., Cunha, G.R., et al. (1999). Roles for Nkx3.1 in prostate development and cancer. *Genes Dev.* 13, 966–977.
- Bowen, C., Stuart, A., Ju, J.H., Tuan, J., Blonder, J., Conrads, T.P., Veenstra, T.D., and Gelmann, E.P. (2007). NKX3.1 homeodomain protein binds to topoisomerase I and enhances its activity. *Cancer Res.* 67, 455–464.
- Champoux, J.J. (2001). DNA topoisomerases: structure, function, and mechanism. *Annu. Rev. Biochem.* 70, 369–413.
- Choi, Y.J., Li, H., Son, M.Y., Wang, X.H., Fornasaglio, J.L., Sobol, R.W., Lee, M., Vijg, J., Imholz, S., Dollé, M.E., et al. (2014). Deletion of individual Ku subunits

in mice causes an NHEJ-independent phenotype potentially by altering apurinic/apyrimidinic site repair. *PLoS ONE* 9, e86358.

Clinckemalie, L., Spans, L., Dubois, V., Laurent, M., Helsen, C., Joniau, S., and Claessens, F. (2013). Androgen regulation of the TMPRSS2 gene and the effect of a SNP in an androgen response element. *Mol. Endocrinol.* 27, 2028–2040.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 322, 1845–1848.

Darzacq, X., Shav-Tal, Y., de Turris, V., Brody, Y., Shenoy, S.M., Phair, R.D., and Singer, R.H. (2007). In vivo dynamics of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* 14, 796–806.

Davis, L., and Maizels, N. (2014). Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. *Proc. Natl. Acad. Sci. USA* 111, E924–E932.

De Santa, F., Barozzi, I., Miettton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.-L., and Natoli, G. (2010). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol.* 8, e1000384.

Dutta, D., Shatalin, K., Epshtein, V., Gottesman, M.E., and Nudler, E. (2011). Linking RNA polymerase backtracking to genome instability in *E. coli*. *Cell* 146, 533–543.

El Hage, A., French, S.L., Beyer, A.L., and Tollervey, D. (2010). Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev.* 24, 1546–1558.

Fong, Y.W., Inouye, C., Yamaguchi, T., Cattoglio, C., Grubisic, I., and Tjian, R. (2011). A DNA repair complex functions as an Oct4/Sox2 coactivator in embryonic stem cells. *Cell* 147, 120–131.

Fong, Y.W., Cattoglio, C., and Tjian, R. (2013). The intertwined roles of transcription and repair proteins. *Mol. Cell* 52, 291–302.

Garcia-Bassets, I., Kwon, Y.S., Telese, F., Prefontaine, G.G., Hutt, K.R., Cheng, C.S., Ju, B.G., Ohgi, K.A., Wang, J., Escoubet-Lozach, L., et al. (2007). Histone methylation-dependent mechanisms impose ligand dependency for gene activation by nuclear receptors. *Cell* 128, 505–518.

Gavrieli, Y., Sherman, Y., and Ben-Sasson, S.A. (1992). Identification of programmed cell death in situ via specific labeling of nuclear DNA fragmentation. *J. Cell Biol.* 119, 493–501.

Gu, J., Lu, H., Tippin, B., Shimazaki, N., Goodman, M.F., and Lieber, M.R. (2007). XRCC4:DNA ligase IV can ligate incompatible DNA ends and can ligate across gaps. *EMBO J.* 26, 1010–1023.

Gurel, B., Ali, T.Z., Montgomery, E.A., Begum, S., Hicks, J., Goggins, M., Eberhart, C.G., Clark, D.P., Bieberich, C.J., Epstein, J.I., and De Marzo, A.M. (2010). NKX3.1 as a marker of prostatic origin in metastatic tumors. *Am. J. Surg. Pathol.* 34, 1097–1105.

Hah, N., Murakami, S., Nagari, A., Danko, C.G., and Kraus, W.L. (2013). Enhancer transcripts mark active estrogen receptor binding sites. *Genome Res.* 23, 1210–1223.

Hamilton, N.K., and Maizels, N. (2010). MRE11 function in response to topoisomerase poisons is independent of its function in double-strand break repair in *Saccharomyces cerevisiae*. *PLoS ONE* 5, e15387.

Herendeen, D.R., Kassavetis, G.A., and Geiduschek, E.P. (1992). A transcriptional enhancer whose function imposes a requirement that proteins track along DNA. *Science* 256, 1298–1303.

Hoeijmakers, J.H. (2009). DNA damage, aging, and cancer. *N. Engl. J. Med.* 361, 1475–1485.

Horoszewicz, J.S., Leong, S.S., Chu, T.M., Wajsman, Z.L., Friedman, M., Papsidero, L., Kim, U., Chai, L.S., Kakati, S., Arya, S.K., and Sandberg, A.A. (1980). The LNCaP cell line—a new model for studies on human prostatic carcinoma. *Prog. Clin. Biol. Res.* 37, 115–132.

Hsieh, C.L., Fei, T., Chen, Y., Li, T., Gao, Y., Wang, X., Sun, T., Sweeney, C.J., Lee, G.S., Chen, S., et al. (2014). Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. *Proc. Natl. Acad. Sci. USA* 111, 7319–7324.

Ju, B.G., Lunyak, V.V., Perissi, V., Garcia-Bassets, I., Rose, D.W., Glass, C.K., and Rosenfeld, M.G. (2006). A topoisomerase II β -mediated dsDNA break required for regulated transcription. *Science* 312, 1798–1802.

Jung, S.Y., Malovannaya, A., Wei, J., O'Malley, B.W., and Qin, J. (2005). Proteomic analysis of steady-state nuclear hormone receptor coactivator complexes. *Mol. Endocrinol.* 19, 2451–2465.

Kaikkonen, M.U., Spann, N.J., Heinz, S., Romanoski, C.E., Allison, K.A., Stender, J.D., Chun, H.B., Tough, D.F., Prinjha, R.K., Benner, C., and Glass, C.K. (2013). Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol. Cell* 51, 310–325.

Kim, T.-K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465, 182–187.

Knudsen, K.E., Arden, K.C., and Cavenee, W.K. (1998). Multiple G1 regulatory elements control the androgen-dependent proliferation of prostatic carcinoma cells. *J. Biol. Chem.* 273, 20213–20222.

Kouzine, F., and Levens, D. (2007). Supercoil-driven DNA structures regulate genetic transactions. *Frontiers in bioscience: a journal and virtual library. Front Biosci.* 12, 4409–4423.

Kouzine, F., Gupta, A., Baranello, L., Wojtowitz, D., Ben-Aissa, K., Liu, J., Przytycka, T.M., and Levens, D. (2013). Transcription-dependent dynamic supercoiling is a short-range genomic force. *Nat. Struct. Mol. Biol.* 20, 396–403.

Kretzschmar, M., Meisterernst, M., and Roeder, R.G. (1993). Identification of human DNA topoisomerase I as a cofactor for activator-dependent transcription by RNA polymerase II. *Proc. Natl. Acad. Sci. USA* 90, 11508–11512.

Kuzminov, A. (2001). Single-strand interruptions in replicating chromosomes cause double-strand breaks. *Proc. Natl. Acad. Sci. USA* 98, 8241–8246.

Kwak, H., Fuda, N.J., Core, L.J., and Lis, J.T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* 339, 950–953.

Lai, F., Orom, U.A., Cesaroni, M., Beringer, M., Taatjes, D.J., Blobel, G.A., and Shiekhattar, R. (2013). Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* 494, 497–501.

Laine, J.P., Opresko, P.L., Indig, F.E., Harrigan, J.A., von Kobbe, C., and Bohr, V.A. (2003). Werner protein stimulates topoisomerase I DNA relaxation activity. *Cancer Res.* 63, 7136–7146.

Lam, M.T., Cho, H., Lesch, H.P., Gosselin, D., Heinz, S., Tanaka-Oishi, Y., Benner, C., Kaikkonen, M.U., Kim, A.S., Kosaka, M., et al. (2013). Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature* 498, 511–515.

Lam, M.T., Li, W., Rosenfeld, M.G., and Glass, C.K. (2014). Enhancer RNAs and regulated transcriptional programs. *Trends Biochem. Sci.* 39, 170–182.

Le May, N., Fradin, D., Iltis, I., Bougnères, P., and Egly, J.M. (2012). XPG and XPF endonucleases trigger chromatin looping and DNA demethylation for accurate expression of activated genes. *Mol. Cell* 47, 622–632.

Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X., et al. (2013a). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* 498, 516–520.

Li, H., Marple, T., and Hasty, P. (2013b). Ku80-deleted cells are defective at base excision repair. *Mutat. Res.* 745–746, 16–25.

Liu, L.F., and Wang, J.C. (1987). Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. USA* 84, 7024–7027.

Ma, J., and Wang, M. (2014). Interplay between DNA supercoiling and transcription elongation. *Transcription* 5. Published online April 7, 2014.

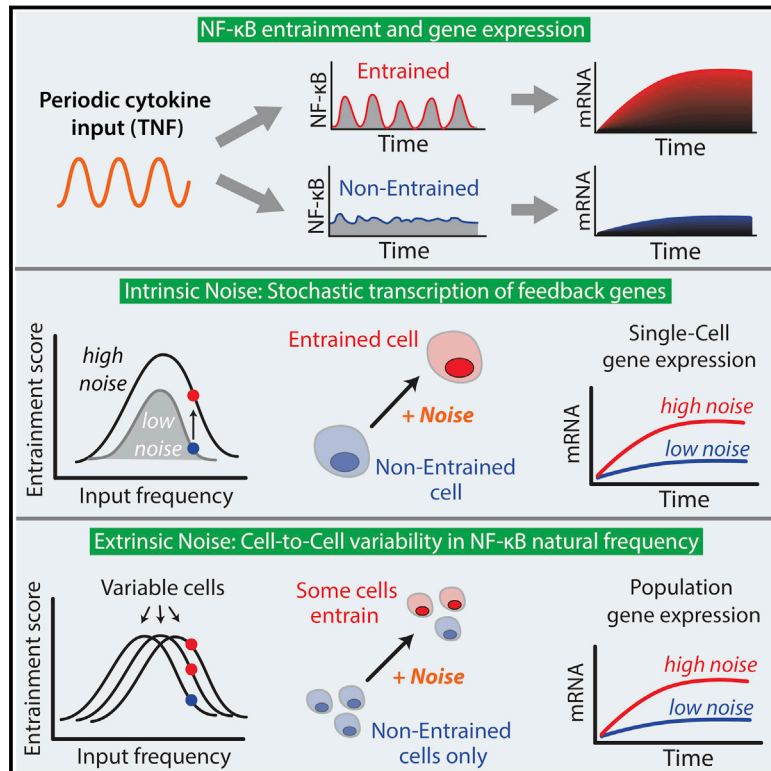
Madden, K.R., and Champoux, J.J. (1992). Overexpression of human topoisomerase I in baby hamster kidney cells: hypersensitivity of clonal isolates to camptothecin. *Cancer Res.* 52, 525–532.

Maldonado, E., Shiekhattar, R., Sheldon, M., Cho, H., Drapkin, R., Rickert, P., Lees, E., Anderson, C.W., Linn, S., and Reinberg, D. (1996). A human RNA

- polymerase II complex associated with SRB and DNA-repair proteins. *Nature* 381, 86–89.
- Mayeur, G.L., Kung, W.J., Martinez, A., Izumiya, C., Chen, D.J., and Kung, H.J. (2005). Ku is a novel transcriptional recycling coactivator of the androgen receptor in prostate cancer cells. *J. Biol. Chem.* 280, 10827–10833.
- Melgar, M.F., Collins, F.S., and Sethupathy, P. (2011). Discovery of active enhancers through bidirectional expression of short transcripts. *Genome Biol.* 12, R113.
- Melo, C.A., Drost, J., Wijchers, P.J., van de Werken, H., de Wit, E., Oude Vrielink, J.A., Elkon, R., Melo, S.A., Léveillé, N., Kalluri, R., et al. (2013). eRNAs are required for p53-dependent enhancer activity and gene transcription. *Mol. Cell* 49, 524–535.
- Merino, A., Madden, K.R., Lane, W.S., Champoux, J.J., and Reinberg, D. (1993). DNA topoisomerase I is involved in both repression and activation of transcription. *Nature* 365, 227–232.
- Mo, X., and Dynan, W.S. (2002). Subnuclear localization of Ku protein: functional association with RNA polymerase II elongation sites. *Mol. Cell Biol.* 22, 8088–8099.
- Mousavi, K., Zare, H., Dell'orso, S., Grontved, L., Gutierrez-Cruz, G., Derfoul, A., Hager, G.L., and Sartorelli, V. (2013). eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. *Mol. Cell* 51, 606–617.
- Natoli, G., and Andrau, J.-C. (2012). Noncoding transcription at enhancers: general principles and functional models. *Annu. Rev. Genet.* 46, 1–19.
- Nimonkar, A.V., Genschel, J., Kinoshita, E., Polaczek, P., Campbell, J.L., Wyman, C., Modrich, P., and Kowalczykowski, S.C. (2011). BLM-DNA2-RPA-MRN and EXO1-BLM-RPA-MRN constitute two DNA end resection machineries for human DNA break repair. *Genes Dev.* 25, 350–362.
- Pedersen, J.M., Fredsoe, J., Roedgaard, M., Andreasen, L., Mundbjerg, K., Kruhoffer, M., Brinch, M., Schierup, M.H., Bjergbaek, L., and Andersen, A.H. (2012). DNA topoisomerases maintain promoters in a state competent for transcriptional activation in *Saccharomyces cerevisiae*. *PLoS Genet.* 8, e1003128.
- Pennacchio, L.A., Bickmore, W., Dean, A., Nobrega, M.A., and Bejerano, G. (2013). Enhancers: five essential questions. *Nat. Rev. Genet.* 14, 288–295.
- Polkinghorn, W.R., Parker, J.S., Lee, M.X., Kass, E.M., Spratt, D.E., Iaquinta, P.J., Arora, V.K., Yen, W.F., Cai, L., Zheng, D., et al. (2013). Androgen receptor signaling regulates DNA repair in prostate cancers. *Cancer Discov.* 3, 1245–1253.
- Pommier, Y., Pourquier, P., Fan, Y., and Strumberg, D. (1998). Mechanism of action of eukaryotic DNA topoisomerase I and drugs targeted to the enzyme. *Biochim. Biophys. Acta* 1400, 83–105.
- Pommier, Y., Barcelo, J.M., Rao, V.A., Sordet, O., Jobson, A.G., Thibaut, L., Miao, Z.H., Seiler, J.A., Zhang, H., Marchand, C., et al. (2006). Repair of topoisomerase I-mediated DNA damage. *Prog. Nucleic Acid Res. Mol. Biol.* 81, 179–229.
- Pommier, Y., Leo, E., Zhang, H., and Marchand, C. (2010). DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. *Chem. Biol.* 17, 421–433.
- Potenski, C.J., Niu, H., Sung, P., and Klein, H.L. (2014). Avoidance of ribonucleotide-induced mutations by RNase H2 and Srs2-Exo1 mechanisms. *Nature* 511, 251–254.
- Price, B.D., and D'Andrea, A.D. (2013). Chromatin remodeling at DNA double-strand breaks. *Cell* 152, 1344–1354.
- Sacho, E.J., and Maizels, N. (2011). DNA repair factor MRE11/RAD50 cleaves 3'-phosphotyrosyl bonds and resects DNA to repair damage caused by topoisomerase 1 poisons. *J. Biol. Chem.* 286, 44945–44951.
- Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* 15, 272–286.
- Shykind, B.M., Kim, J., Stewart, L., Champoux, J.J., and Sharp, P.A. (1997). Topoisomerase I enhances TFIIID-TFIIA complex assembly during activation of transcription. *Genes Dev.* 11, 397–407.
- Simmons, D.T., Roy, R., Chen, L., Gai, D., and Trowbridge, P.W. (1998). The activity of topoisomerase I is modulated by large T antigen during unwinding of the SV40 origin. *J. Biol. Chem.* 273, 20390–20396.
- Sims, D., Sudbery, I., Iltott, N.E., Heger, A., and Ponting, C.P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. *Nat. Rev. Genet.* 15, 121–132.
- Stewart, L., Redinbo, M.R., Qiu, X., Hol, W.G., and Champoux, J.J. (1998). A model for the mechanism of human topoisomerase I. *Science* 279, 1534–1541.
- Stracker, T.H., and Petrini, J.H. (2011). The MRE11 complex: starting from the ends. *Nat. Rev. Mol. Cell Biol.* 12, 90–103.
- Tan, P.Y., Chang, C.W., Chng, K.R., Wansa, K.D., Sung, W.K., and Cheung, E. (2012). Integration of regulatory networks by NKX3-1 promotes androgen-dependent prostate cancer survival. *Mol. Cell Biol.* 32, 399–414.
- Tsao, Y.P., Russo, A., Nyamuswa, G., Silber, R., and Liu, L.F. (1993). Interaction between replication forks and topoisomerase I-DNA cleavable complexes: studies in a cell-free SV40 DNA replication system. *Cancer Res.* 53, 5908–5914.
- Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., Qiu, J., Liu, W., Kaikkonen, M.U., Ohgi, K.A., et al. (2011). Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* 474, 390–394.
- Wimberly, H., Shee, C., Thornton, P.C., Sivaramakrishnan, P., Rosenberg, S.M., and Hastings, P.J. (2013). R-loops and nicks initiate DNA breakage and genome instability in non-growing *Escherichia coli*. *Nat. Commun.* 4, 2115.

Noise Facilitates Transcriptional Control under Dynamic Inputs

Graphical Abstract



Authors

Ryan A. Kellogg, Savaş Tay

Correspondence

savas.tay@bsse.ethz.ch

In Brief

Intrinsic biochemical noise in individual cells, traditionally considered as harmful to signal transduction, improves NF-κB oscillation and entrainment, whereas cell-to-cell variability in NF-κB natural frequency creates population robustness. Together, the two types of transcriptional noise enable signal entrainment over a wider range of dynamic inputs.

Highlights

- Periodic TNF stimulation entrains and amplifies NF-κB oscillations
- Amplified NF-κB oscillation increases transcriptional efficiency
- Intrinsic biochemical noise improves oscillation and entrainment of single cells
- Extrinsic noise allows population to entrain robustly under a wider range of inputs



Noise Facilitates Transcriptional Control under Dynamic Inputs

Ryan A. Kellogg¹ and Savaş Tay^{1,*}

¹Department of Biosystems Science and Engineering, ETH Zürich 4058, Switzerland

*Correspondence: savas.tay@bsse.ethz.ch
<http://dx.doi.org/10.1016/j.cell.2015.01.013>

SUMMARY

Cells must respond sensitively to time-varying inputs in complex signaling environments. To understand how signaling networks process dynamic inputs into gene expression outputs and the role of noise in cellular information processing, we studied the immune pathway NF- κ B under periodic cytokine inputs using microfluidic single-cell measurements and stochastic modeling. We find that NF- κ B dynamics in fibroblasts synchronize with oscillating TNF signal and become entrained, leading to significantly increased NF- κ B oscillation amplitude and mRNA output compared to non-entrained response. Simulations show that intrinsic biochemical noise in individual cells improves NF- κ B oscillation and entrainment, whereas cell-to-cell variability in NF- κ B natural frequency creates population robustness, together enabling entrainment over a wider range of dynamic inputs. This wide range is confirmed by experiments where entrained cells were measured under all input periods. These results indicate that synergy between oscillation and noise allows cells to achieve efficient gene expression in dynamically changing signaling environments.

INTRODUCTION

Understanding how cells efficiently process information in rapidly changing and noisy environments is a fundamental problem in biology. Cells experience environments that fluctuate over time during physiological conditions such as inflammation, where oscillating input signals can occur due to pulsatile secretion of signaling molecules from immune cells (Goldbeter et al., 1990; Han et al., 2012), propagating signaling waves (Falcke, 2003; Schütze et al., 2011; Yde et al., 2011), or by coupling between upstream pathways (Gérard and Goldbeter, 2012; Goldbeter and Pourquie, 2008; Yang et al., 2010; Yoshiura et al., 2007). How cells process such dynamic inputs into functional gene expression outputs is not well understood. Further, signaling systems are subject to biochemical noise originating from stochastic molecular interactions, leading to system noise and cell-to-cell variability in response to input signals. While it is a common belief that noise is harmful to information processing (Cheong et al., 2011), cell signaling pathways perform with

remarkable robustness despite ever-present system noise and variability (Little et al., 1999). It is not clear how cell-signaling pathways overcome system noise and whether there are functional roles for noise in cellular information processing.

Signaling systems often employ oscillatory network architecture to process environmental inputs (Levine et al., 2013). For example, specific transcriptional responses can be achieved by encoding the dose or identity of a constant input signal by modulating oscillatory response dynamics (Kupzig et al., 2005). Theoretically, oscillation can be advantageous also in the processing of fluctuating and noisy input signals. For example, dynamic inputs that contain noise can be transmitted efficiently in an oscillating system through a phenomenon called stochastic resonance (Douglass et al., 1993), previously observed in neuronal circuits (McDonnell and Ward, 2011). Nevertheless, such a beneficial role for biochemical system noise in the processing of fluctuating environmental signals has not been shown.

To study how oscillation and system noise may interact in processing of dynamic input signals, we consider the NF- κ B system, a gene regulatory network central to immune functions and many diseases, including autoimmunity and cancer (Hayden and Ghosh, 2008). NF- κ B pathway activation by TNF cytokine leads to oscillations in p65:p50 heterodimer localization between the cytoplasm and nucleus (Hayden and Ghosh, 2008; Hoffmann et al., 2002; Nelson et al., 2004), mediated by NF- κ B-dependent induction of negative feedback genes of the I κ B family (Figure 1A). Pathway activation through IKK under TNF occurs in a digital, switch-like fashion (Tay et al., 2010).

NF- κ B oscillations are subject to intrinsic and extrinsic noise, leading to variable timing between cells that obscures single-cell behavior in population analyses (Swain et al., 2002; Tay et al., 2010). Sources of extrinsic noise include different signaling histories and uneven cell division leading to variation in protein abundance (Huh and Paulsson, 2011). Variation in TNF receptor or NF- κ B molecules create different response characteristics between cells (Tay et al., 2010). Significant contributions to intrinsic (biochemical) noise in NF- κ B include burst-like transcription of I κ B and A20 negative feedback genes, and receptor-ligand interaction at low ligand concentration (Elowitz et al., 2002; Tay et al., 2010). Another negative feedback gene I κ B ϵ is induced with a 45 min delay compared to I κ B α , which is optimally timed for increasing cell-to-cell oscillation variability, suggesting that transcriptional noise might provide a functional advantage (Ashall et al., 2009; Paszek et al., 2010).

The function of NF- κ B oscillation is not fully understood. Other pathways like p53 and Notch convert between oscillatory and

non-oscillatory response to achieve specific cell fate responses (Dolmetsch et al., 1998; Kageyama et al., 2008; Purvis et al., 2012; Purvis and Lahav, 2013). The frequency of oscillation in ERK, Crz1, and NFAT4 is altered depending on input signal concentration, achieving expression control across diverse promoters through frequency modulation (Albeck et al., 2013; Beridge et al., 2003; Cai et al., 2008; Dolmetsch et al., 1998; Eldar and Elowitz, 2010; Shankaran et al., 2009; Yissachar et al., 2013). However, NF- κ B oscillation frequency (90 to 100 min peak-to-peak interval) is unchanged across a wide range of input concentrations (Longo et al., 2013; Tay et al., 2010; Turner et al., 2010), and it is uncertain how NF- κ B oscillation changes and directs gene expression in response to a fluctuating input.

Oscillatory systems can experience resonance, where a periodic stimulus leads to amplified output (Abraham et al., 2010; Pikovsky et al., 2003). Periodic input may also entrain or synchronize a population of oscillators so that all oscillators adopt the same frequency and phase. Entrainment leading to resonant amplification of NF- κ B oscillations may occur for input signals that fluctuate at a rate similar to the NF- κ B natural frequency, increasing the sensitivity of the NF- κ B system especially to small signals. Theoretical studies predict that periodic input to NF- κ B may generate entrainment, quasiperiodic oscillations, or even chaos (Jensen and Krishna, 2012; Wang et al., 2011). Entrainment, with prominent examples from circadian rhythms and brain waves, allows oscillatory signaling and transcriptional pathways to synchronize and work in harmony (Reppert and Weaver, 2002; Varela et al., 2001). It is conceivable that entrainment could reduce cell-to-cell NF- κ B oscillation variability, leading to homogenous transcriptional responses at the population level. Nevertheless, experimental studies are lacking on whether NF- κ B can experience resonance and entrainment, and whether there is impact on gene expression output and variability (Longo et al., 2013; Tay et al., 2010; Turner et al., 2010).

Although noise is detrimental to signal transmission in linear systems, it can facilitate information transfer in a non-linear system by decreasing the amplitude of a periodic input needed to achieve coupling (Collins et al., 1996; Lindner et al., 2004; Mori and Kai, 2002; Zhou et al., 2002). For example, input noise can facilitate sensory neuron processing (McDonnell and Ward, 2011) and intrinsic noise may cause oscillations to become more robust to perturbation (Paszek et al., 2010; Perc and Marhl, 2003; Vilar et al., 2002). Extrinsic noise (i.e., variation in signaling parameters between cells) may also impact entrainment due to increased population diversity, similar to bacterial bet hedging when external conditions change (Mondragón-Palomino et al., 2011; Süel et al., 2006; Wakamoto et al., 2013). However, it is not known how noise could affect entrainment of a complex and physiological mammalian system such as NF- κ B.

To probe how oscillation and noise together determine NF- κ B-dependent transcription in dynamic settings, we used a microfluidics-based experimental pipeline that enabled automated cell stimulation, live imaging, and gene expression measurements (Gómez-Sjöberg et al., 2007; Junkin and Tay, 2014; Kellogg et al., 2014; Tay et al., 2010). We delivered various TNF cytokine inputs to p65^{-/-} mouse 3T3 fibroblast cells expressing p65/DsRed fusion protein at near wild-type levels (Lee et al.,

2009; Tay et al., 2010) (Figures 1B and 2A and 3A). The microfluidic chip utilizes computer controlled PDMS membrane valves, allowing constant perfusion or periodic pulsing of signaling factors. Ninety-six independent cell culture experiments each with complex fluidic conditions can be maintained in parallel. Cell images acquired at 5 min intervals were automatically analyzed, extracting thousands of single-cell trajectories of NF- κ B nuclear intensity over time (Kellogg et al., 2014). Following periodic TNF stimulation, cells were retrieved for gene expression analysis in a high-throughput microfluidic qPCR system to understand the influence of entrainment on target gene expression (Kellogg et al., 2014). Furthermore, we performed stochastic simulations using an established model of NF- κ B (Tay et al., 2010) and varied both intrinsic and extrinsic noise to interpret our experimental findings and understand the role of noise and oscillations for NF- κ B dynamic signal processing.

RESULTS

Constant TNF Stimulation Generates Sustained, Noisy NF- κ B Oscillations

Previous studies of NF- κ B dynamics were subject to TNF ligand loss due to degradation and cellular internalization, leading to damped oscillations (Tay et al., 2010). Here, constant TNF concentration was achieved by perfusion of fresh TNF-containing media using an on-chip peristaltic pump (Figure 1B). Under constant TNF concentration, we observed NF- κ B oscillations sustaining longer than 24 hr with mean period approximately 90 min (Figures 1C and 1D and Movie S1). Cells exhibited different natural frequencies (between-cell variability) and cycle-to-cycle timing fluctuation (within-cell variability) (Figure 1F). NF- κ B oscillation is robust to changes in dose, and lowered dose, which generates high receptor-ligand noise, modestly lengthened the average period and increased period variability (CV: coefficient of variation, in Figure 1E). These findings show that NF- κ B oscillation sustains under constant TNF input, pointing to a conserved function for oscillations.

To understand how noise underlies oscillation variability, we performed simulations of NF- κ B dynamics under constant TNF concentration. We used a stochastic single-cell model, which faithfully reproduces the NF- κ B dynamics in single 3T3 fibroblast cells used in this study (Tay et al., 2010). The simulations showed sustained oscillation and period characteristics similar to our experiments (Figure 1G) (Lipniacki et al., 2004; Lipniacki et al., 2007; Tay et al., 2010). Varying intrinsic noise in the model associated with changes in cycle-to-cycle variability, while changing extrinsic noise affected variability in average (natural) oscillation period between cells (Figures 1G and 1H). Magnitudes of variability for simulations matched that of experimental measurements, supporting an appropriate balance between intrinsic and extrinsic noise in the model (Figure 1I).

Single-Cell NF- κ B Dynamics becomes Entrained under an Oscillating Cytokine Signal

During inflammation cells operate under dynamic TNF signals, which may interfere with NF- κ B oscillations needed for processing of input dose information and differential gene expression (Tay et al., 2010). Depending on frequency and amplitude,

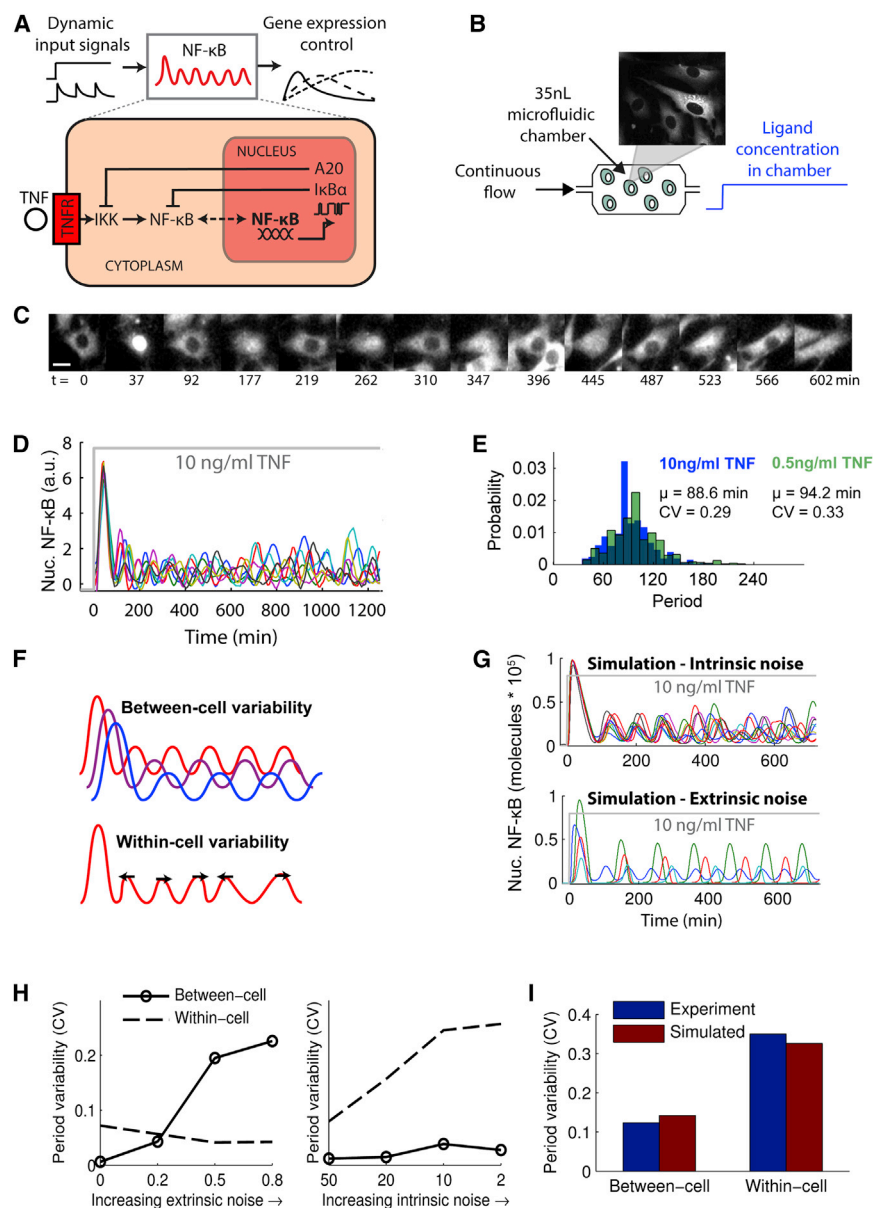


Figure 1. Noise Origins of Sustained NF-κB Oscillation and Heterogeneity

(A) NF-κB transcription factor oscillates between cytoplasm and nucleus in response to inflammatory signals. NF-κB dynamics relay external signals to gene expression outputs.

(B) We deliver continuous or periodic inputs to cells using microfluidic cell culture. In continuous mode, TNF is flowed over cells to maintain constant concentration.

(C) We record single-cell NF-κB translocation using live-cell microscopy. Images show nucleus-cytoplasm oscillations in NF-κB (p65-dsRed) under continuous TNF perfusion. Scale bar, 10 μm.

(D) Under constant 10 ng/ml TNF concentration, NF-κB shows long sustaining, asynchronous oscillations.

(E) NF-κB oscillates with mean period ~ 90 min under constant high and low dose input ($n = 40$ cells).

(F) Pictorial depiction of within- and between-cell oscillation variability. While cells may have different mean periods (between-cell variability), each cell also exhibits fluctuation in its own oscillation (within-cell variability).

(G) Simulated single cell trajectories show that extrinsic noise increases between-cell variability, and intrinsic noise increases within-cell variability.

(H) In simulations increasing extrinsic noise increases between-cell variability, while increasing intrinsic noise increases within-cell variability.

(I) Experimentally, we observe $\sim 12\%$ period fluctuation between different cells and 34% within the same cells under 10 ng/ml TNF. Variability in simulated traces agrees with experimentally measured values.

periodic input to an oscillator like NF-κB can either entrain or disrupt the oscillation. Entrainment describes when the oscillator becomes phase-locked and synchronized with the driving stimuli (Pikovsky et al., 2003), with prominent examples in biology from circadian rhythms (Leloup and Goldbeter, 2003; Reppert and Weaver, 2002) and brain waves, to synthetic bacterial oscillators (Mondragón-Palomino et al., 2011). On the other hand, when entrainment cannot occur due a significant frequency mismatch between the oscillator and input signal, the result is a disrupted oscillation that is quasiperiodic or even chaotic (Jensen and Krishna, 2012; Pikovsky et al., 2003). How NF-κB responds to sustained periodic inputs that could entrain NF-κB oscillations has not been experimentally investigated so far.

To test the entrainment capacity of NF-κB and how oscillation contributes to gene expression control under fluctuating cyto-

kine signals, we applied TNF inputs to fibroblasts using two stimulation periods (Figure 2A): in the first case, TNF stimulus is applied every 120 min, which indeed efficiently entrained NF-κB after a transient (Figure 2B). In the second case, TNF stimulus was provided every 60 min, which was sufficiently mismatched from the ~ 90 min NF-κB natural period to induce a disrupted, non-entrained NF-κB response in most cells (Figure 2C). Movies S2, S3, and S4 show single-cells under these inputs as well as under 90 min input, and Figure S1 shows the difference between entrained cells and cells oscillating under constant TNF signal. Images selected at three time points show that for 60 min input NF-κB oscillations remain asynchronous in the population, and for 120 min input NF-κB oscillates synchronously across cells (Figures 2D and 2E, also see and Movies S2, S3, and S4). Population phase variability, a measure of synchrony, remains high during the time course with 60 min stimulation, but it quickly reduces during 120 min stimulation as the population entrains (Figures 2F and S1B). Simulations with our comprehensive NF-κB model also reproduced non-entrained and entrained responses under similar inputs (Figure 2G).

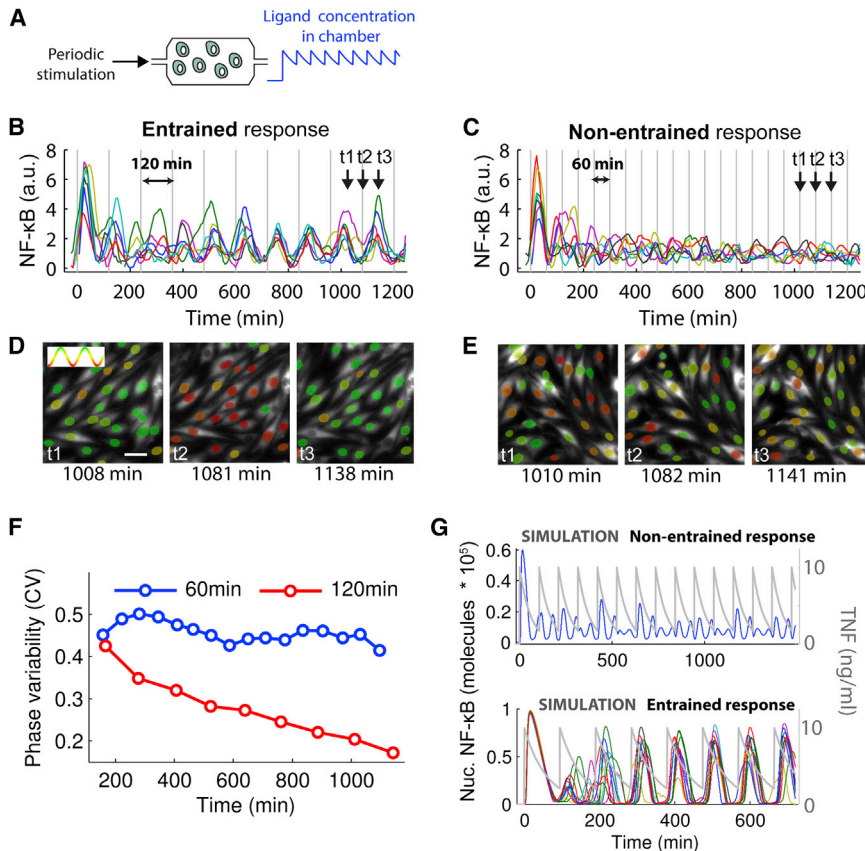


Figure 2. Periodic TNF Stimulation Can Entrain or Disrupt NF-κB Oscillations

(A) We deliver periodic inputs to cells using a microfluidic cell culture chip. In periodic mode, TNF is replaced at specified intervals, and ligand decay (due to cell uptake and degradation) leads to a periodic sawtooth concentration profile.

(B) Single-cell traces for stimulation at 120 min input period, which entrains and synchronizes NF-κB oscillations.

(C) Single-cell NF-κB trajectories measured for stimulation at 60 min input, which disrupts NF-κB oscillations.

(D) Image time-series for 120 min periodic input for times t_1 – t_3 indicated by arrows in B. The entrained cell population oscillates synchronously. Inset: Nucleus color indicates nuclear NF-κB intensity from red (low) to green (high). Scale bar, 25 μ m.

(E) Image time series for 60 min periodic input for times t_1 – t_3 indicated by arrows in (C). The cell population does not synchronize.

(F) Phase variability for 60 min stimulation remains constant over time. In contrast, during 120 min stimulation (red), phase variability decreases as the cell population synchronizes over time.

(G) Simulations reproduce non-entrained and entrained responses. See also Figure S1.

Entrained NF-κB Oscillation Improves Gene Expression Efficiency and Reduces Cell-to-Cell Variability

NF-κB regulates hundreds of pro- and anti-inflammatory genes (Hao and Baltimore, 2013). To understand the influence of entrained versus disorderly NF-κB dynamics in gene expression, we measured time-dependent expression of target genes for 120 min and 60 min periodic TNF stimulation using microfluidic qPCR (Figures 3A and 3B, S2, and S3). Cells stimulated in independent chambers of the cell culture chip were harvested for expression analysis at 30 min time increments (Figure 3A) (See protocols in Kellogg et al., 2014). Under the entraining 120 min input, gene expression output is notably enhanced, especially in genes with later induction times (Figure 3C, red lines). In contrast, 60 min input that leads to non-entrained NF-κB response caused an impaired transcriptional response (Figure 3C, blue lines). Importantly, the difference in measured mRNA expression is not due to a difference in total TNF exposure, as there is greater TNF exposure for 60 min input (Figure 3D).

We analyzed single-cell NF-κB trajectories for entrained and non-entrained conditions to identify what might give rise to the observed gene expression difference. Since differences are most evident in the later part of the time course, we focused our analysis to time after 500 min. We measured NF-κB area under the curve (AUC) and determined the extent that each trace is oscillatory versus non-oscillatory by power spectral analysis. The AUC is a measure of total NF-κB protein localization into the nucleus, which did not change in a significant way to explain

the observed gene expression difference. Entrained compared to non-entrained cells showed only 9% increase in NF-κB area. However, we measured 83% increase in NF-κB oscillatory energy (Figure 3E). Non-entraining input at 60 min mostly resulted in non-oscillatory localization profiles, while entraining input at 120 min resulted in strong oscillations with large amplitude (Figure 3F).

To understand how increased oscillation magnitude under entraining input could lead to higher transcriptional output, we simulated traces for 60 and 120 min sawtooth input, similar to those used in experiments. Simulated NF-κB single cells exhibited similar changes in area and oscillatory energy for entrained versus non-entrained conditions (Figures 3G and S2). However, the existing transcriptional model, which assumed that NF-κB binding to DNA increases linearly with nuclear NF-κB concentration (Tay et al., 2010), did not reproduce increased gene expression for the entrained condition (Figure S2). Experiments indicate that NF-κB binds DNA cooperatively with Hill coefficient ~ 4 (Phelps et al., 2000) (Figure S2A). Introducing this non-linearity in our model created significantly increased transcriptional output for entrained versus non-entrained conditions, in agreement with our experiments (Figures 3G and S2C) (Wee et al., 2012). Increasing intrinsic noise led to stronger oscillations and further amplified the NF-κB-induced gene expression (Figure 3G). Thus, entraining input leads to strengthening of NF-κB oscillations, which are further amplified by noise to drive increased transcriptional output.

We asked whether entrainment could reduce cell-to-cell variability in transcriptional output in our simulations. Comparing the coefficient of variation of mRNA output over time indicates that

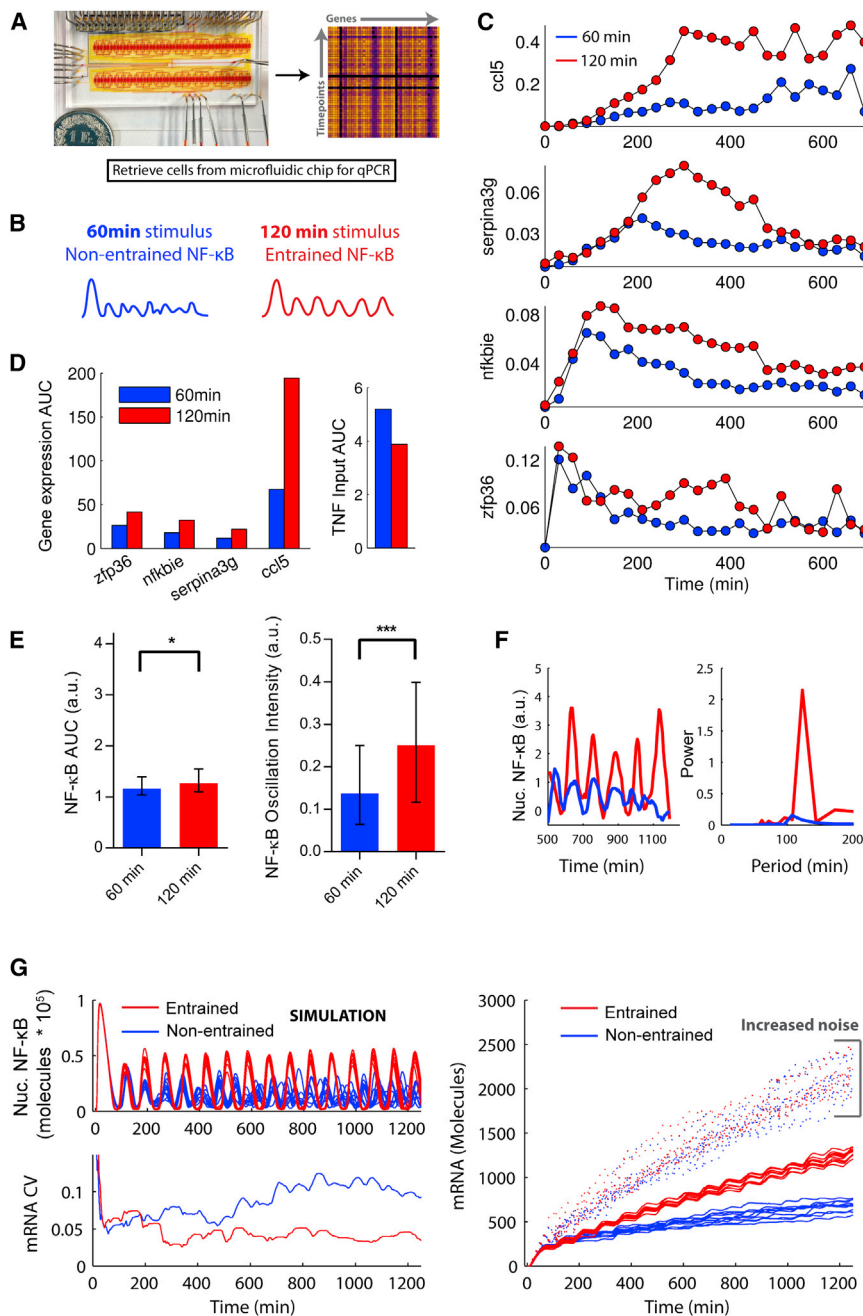


Figure 3. Entrained NF-κB Oscillations Improve Transcriptional Efficiency

(A) Cells are cultured and provided periodic stimulation on chip and harvested for qPCR analysis. (B) TNF stimulation with 60 min period (blue) leads to non-entrained NF-κB response, and most individual cells do not synchronize with the input. (C) NF-κB regulated gene expression under non-entrained (60 min stimulation) and entrained (120 min stimulation) conditions. Higher transcriptional output is seen when NF-κB oscillations are entrained. Enhanced transcription occurs consistently for early, middle, and late genes. The effect is most pronounced for late responding genes (i.e., *ccl5*). (D) Gene expression output measured by area under curve (AUC). AUC is higher for entrained compared to non-entrained NF-κB response. Although 120 min stimulation increases transcript production, it is not due to higher TNF exposure, which is lower compared to 60 min stimulation. (E) Analysis of single-cell NF-κB trajectories shows modest increase in response area ($p = 0.04$) and strong increase in oscillation energy ($p = 0.0002$) (bars indicate median \pm interquartile range, p values by Mann-Whitney test.) (F) Example NF-κB trajectories (for later part of time course starting at 500 min) and corresponding power spectra for 60 and 120 min input (blue and red lines, respectively), showing stronger oscillation under entrained (120 min) input. (G) Stochastic NF-κB simulation of cells under either entraining or non-entraining input (left) and gene expression output (right). Due to non-linear binding of NF-κB to DNA, stronger oscillation under entraining input creates increased gene expression output, in agreement with experiments. Increasing intrinsic noise amplifies oscillations and leads to even higher transcription output. mRNA cell-to-cell variability (measured by Coefficient of Variation, CV) is lower for entrained cells, indicating that entrainment reduces cell-to-cell mRNA variability compared to non-entrained cells. See also Figure S2 and S3.

cell-to-cell transcription variability is significantly reduced under entraining input (Figure 3G). This result is consistent across simulated early, middle, and late-response genes (Figure S3). With reduced mRNA variability between individual cells, oscillations appear even in the population averaged experimental time course, especially in late genes (Figure 3C). Therefore, through entrainment that reduces gene regulatory and gene expression variability between cells, one may increase response homogeneity of a cell population.

These results reveal the important role for NF-κB oscillations in generating efficient transcription and indicate that periodic

show reduced cell-cell variability in mRNA level for entraining input. Therefore entrainment provides a way to increase both expression output and homogeneity of a cell population.

Stochastic Modeling Shows Noise-Enhanced NF-κB Oscillation and Entrainment

We next turned to simulations to evaluate the robustness of NF-κB entrainment to changes in the TNF input period and the influence of noise. Using the deterministic implementation of our model, we simulated periodic TNF stimulation of NF-κB in single cells and calculated entrainment ranges. In the space spanned

signaling inputs can amplify transcriptional outputs by resonantly stimulating oscillatory pathways like NF-κB, with a beneficial role for intrinsic noise in further improving the transcriptional output. Moreover, simulations and experiments

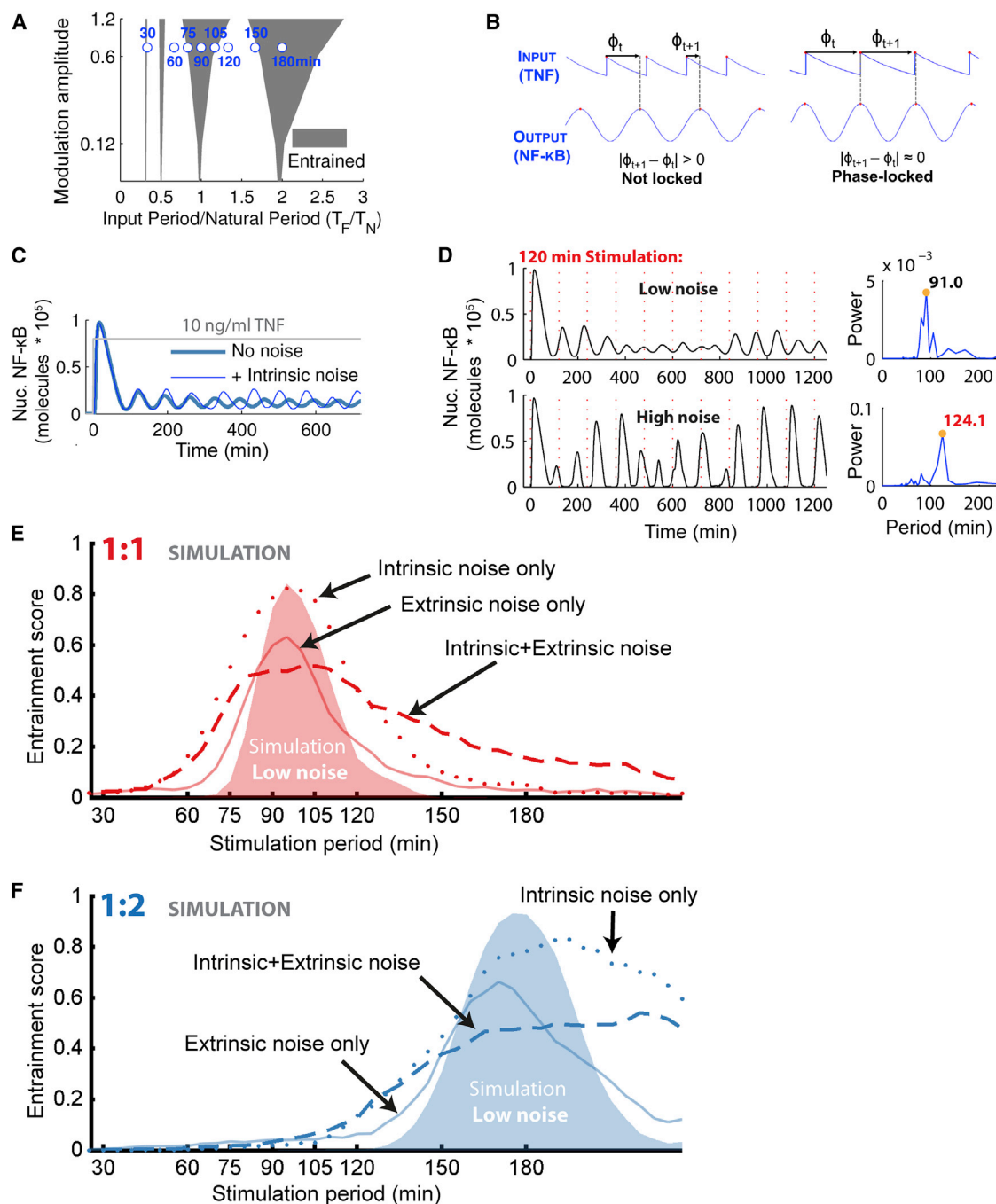


Figure 4. Stochastic Modeling Predicts Entrainment to be Robust and that Noise Underlies Enhancement in Oscillation and Entrainment Range

(A) Deterministic Arnold tongues (gray shaded regions) computed for decay-type TNF input show that entrainment is readily achieved in narrow regions around 90 min and 180 min ($T_F/T_N = 1, 2$) periodic stimulation (10 ng/ml TNF). Entrainment is also possible for 30 and 45 min input ($T_F/T_N = 1/3, 1/2$). Locations of experimentally tested values are indicated by blue circles.

(B) Input-output phase relationship and phase-locking. Phase between TNF input and NF-κB output is calculated as the distance from each NF-κB peak to the start of the previous TNF cycle, normalized by the input period. When phase change between cycles is less than a threshold ($|\phi_{t+1} - \phi_t| < 0.15$), input and output are considered phase-locked. Locking can occur at 1:1 ratio (one input cycle for one output cycle), or other ratios such as 1:2 (one input cycle for two output cycles).

(C) Adding intrinsic noise amplifies and sustains NF-κB oscillations in the model under constant TNF input (example single cell traces are shown).

(D) Comparison of entrainment in simulated NF-κB trajectories under low and high intrinsic noise. Under 120 min periodic TNF stimulation, high noise leads to an entrained response indicated by 120 min peak in the power spectrum. In contrast, the response under low noise is not entrained as seen in the power spectrum that shows weaker oscillations and only at the natural period.

(legend continued on next page)

by input modulation amplitude and period (T_F), entrainment occurs in triangular regions called Arnold Tongues (Figure 4A) (Erzberger et al., 2013; Jensen and Krishna, 2012). On the edges of Arnold Tongues synchrony between the input and oscillator breaks down leading to quasiperiodic or aperiodic rhythms. Deterministic Arnold tongues for NF- κ B indicated entrainment principally when stimulation period is near 1:1 or 1:2 ratio with the natural period and ($T_F/T_N = 1, 2$) (Figure 4A), meaning that entrainment is expected when the stimulation occurs with a period near 90 min or near 180 min under 10 ng/ml TNF input.

To understand the role of different noise levels in entrainment, we simulated periodic TNF signals with varied intrinsic and extrinsic noise conditions and quantified NF- κ B phase locking by comparing phase of the next cycle ϕ_{t+1} to that of the current cycle ϕ_t . If the phase difference $|\phi_{t+1} - \phi_t|$ is less than a threshold (0.15) then the response was considered locked over that cycle (Figure 4B). Our hybrid model based on Gillespie algorithm incorporates experimentally verified intrinsic noise in TNF receptor-ligand binding, which is dominant at small TNF doses, and in transcription of I κ B α and A20 that constitute the main negative-feedback loops leading to oscillations (Tay et al., 2010). Particularly, transcriptional noise arises from stochastic interaction of NF- κ B transcription factors with the two copies of I κ B α and A20 genes. We reduced the transcriptional noise by increasing the gene copy number and proportionally reducing gene expression rate per copy to maintain unchanged gene expression and similar NF- κ B natural oscillation period between models. With the stochastic model with greatly reduced intrinsic noise, entrainment occurs for narrow regions around 90 and 180 min stimulation under high dose (10 ng/ml) TNF periodic input (Figures 4E and 4F), similar to those in the deterministic simulations. Simulations with high intrinsic noise under the same TNF input led to a significant broadening of the entrainment regions (Figures 4E and 4F, dotted line). The intrinsic noise level in these simulations was matched to the experimental level in Figure 1. High intrinsic noise in our simulations increased NF- κ B oscillation amplitude of single cells and supported sustained oscillations needed for entrainment (Figure 4C). The power spectrum provides information about entrainment, and degree of entrainment is indicated by the relative amount of spectral power at the input period. Example simulated NF- κ B single-cell trajectories for 120 min input are seen in Figure 4D, showing significantly increased oscillation and spectral power at the input period for the high noise case. To determine how noise effects depend on TNF dose and modulation level, we simulated the same model using computationally efficient stochastic differential equations, which showed that intrinsic noise improves NF- κ B power at input period when the input modulation is smaller (i.e., weaker driving stimuli), as in higher-dose periodic TNF stimulation (Figure S4).

Extrinsic noise generates cell-to-cell variability in NF- κ B natural period (Figures 1G and 1H). When the natural period in an individual cell is sufficiently close to the TNF input period,

entrainment will occur. Extrinsic noise in the system thus increases the probability that at least a portion of cells in the population will entrain to a given input. When we included extrinsic noise in our simulations, we observed a further broadening of entrainment ranges NF- κ B (Figures 4E and 4F). Overall, these simulations indicate that extrinsic and intrinsic noise together enable cells to entrain and drive efficient transcriptional responses for a wider range of dynamical inputs.

NF- κ B Entrainment Range Is Very Broad as Predicted by Noisy Simulations

To experimentally test the robustness of NF- κ B entrainment to changes in the input, we applied TNF inputs with 30 to 180 min periods to fibroblasts cultured in separate chambers of the microfluidic system, under three different TNF doses of 10, 0.5, and 0.1 ng/ml (Figures 5 and S5). The dataset contains analysis of approximately 2,000 cells over 24hrs duration measured every 5 min, creating more than half a million data points (Movies S2, S3, and S4). Heatmaps with one row for each single-cell NF- κ B trajectory show population synchrony that improves with time (Figure 5A). Periodic stimulation with reduced dose leads to even better entrainment (Figure 5A). The fraction of NF- κ B cycles locking to different entrainment ratios was computed for each stimulation condition, and as anticipated 1:1 locking is maximized when then stimulation period is near 90 min and 1:2 locking is maximized for 180 min stimulation (Figure 5B). Surprisingly, we observed cells having entrained oscillations in every input period tested, even in those inputs like 120 min that are not predicted by the deterministic or low noise simulations. Good agreement is seen between experimental entrainment values and high-noise model simulations (both under 10 ng/ml TNF dose) incorporating both extrinsic and intrinsic noise (Figure 5B). We did not observe dependence on cell density (Figure S6).

A consequence of natural period diversity is entrainment heterogeneity, including the ability for different cells in the population to entrain at different ratios. Cells entrained at multiple ratios or did not entrain and exhibited quasiperiodic oscillation (Figure 6C). Period probability follows a multimodal distribution, indicating simultaneous mixture of for example 1:1 and 1:2 locking responses in the population (Figure 6A). Simulations incorporating only extrinsic noise also generate mixed locking responses, indicating that locking heterogeneity can arise from extrinsic noise (Figure 6D).

Period distributions show narrowing with reduced TNF dose, supporting more effective entrainment (Figure 6B). Comparing mean pairwise Spearman correlation for population responses at each input revealed increased correlation as dose decreases and at larger input periods (Figure 6E). Therefore, NF- κ B is more amenable to entrainment for input level in middle of its dose dynamic range (Tay et al., 2010) in agreement with simulation (Figure S5C). Entrainment is more efficient under input periods larger than 60 min, and frequencies larger than 0.02 min^{-1} (50 min

(E and F) Entrainment simulation: Under low-noise simulation (extrinsic noise off, intrinsic noise reduced), entrainment occurs for narrow regions around 90 min stimulation period (1:1 entrainment, E) and around 180 min period (1:2 entrainment, F). Increasing intrinsic noise in the model broadens regions of entrainment (dotted line). Extrinsic noise alone also increases entrainment range for the population. Adding both extrinsic and intrinsic noise further expands entrainment. See also Figure S4.

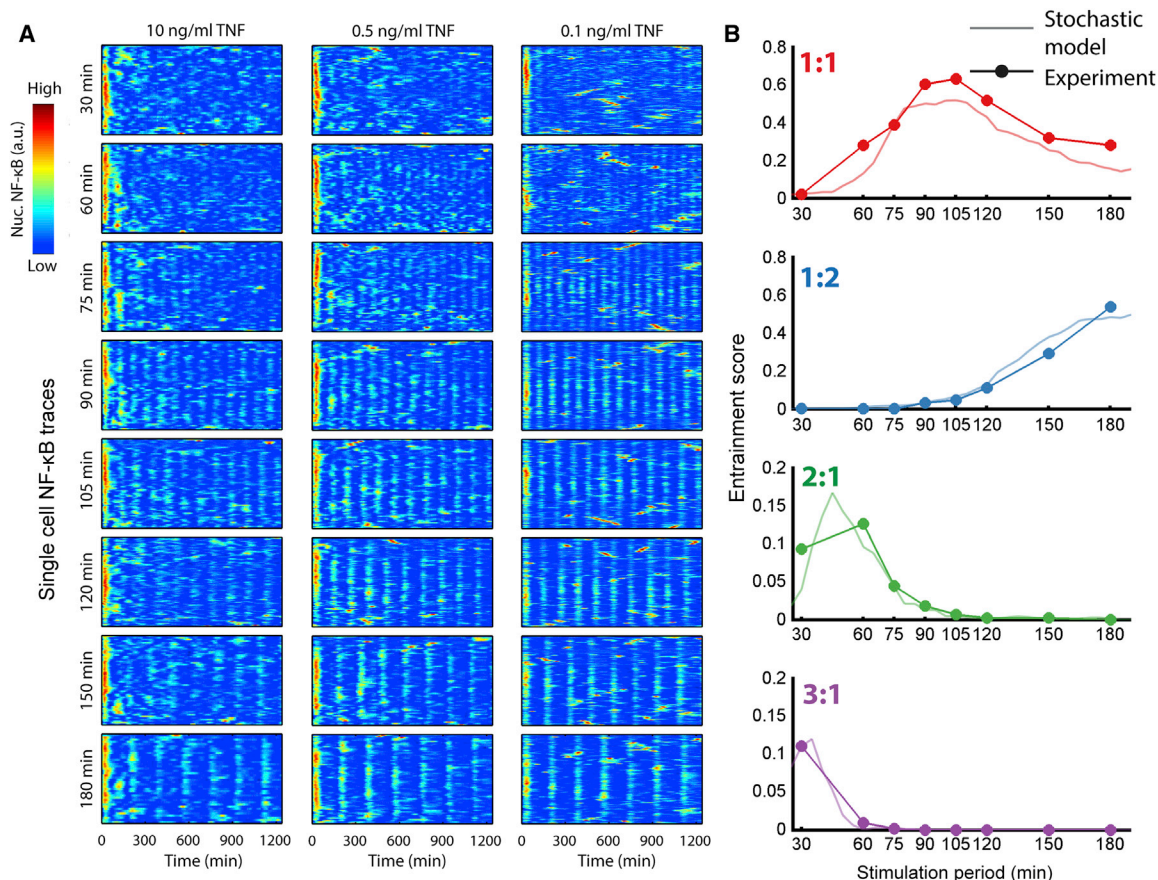


Figure 5. NF-κB Entrainment Range Is Wide and Agrees with Noisy Model Predictions

(A) Heatmaps of single-cell NF-κB trajectories under different doses of TNF (10, 0.5, and 0.1 ng/ml) for stimulation periods ranging from 30 to 180 min. Color indicates NF-κB intensity from low (blue) to high (red). Entrainment of individual cells can be visualized with the appearance of well-aligned peaks. Entrainment and synchronization is pronounced for 90 and 180 min stimulation. Reduced dose leads to greater oscillation synchrony and improved entrainment across all stimulation periods (Figure S5B).

(B) Comparison of entrainment scores for 10 ng/ml stimulation in various locking modes shows agreement between experiments and noisy model prediction. See also Figure S5.

period) are not observed in the single cell power spectra, indicating that NF-κB system acts like a filter that prevents transmittance of rapid TNF input fluctuations into transcription.

DISCUSSION

Here, we provide insight into the function of transcription factor dynamics and noise in gene expression control under fluctuating signaling inputs. Sustained, heterogeneous single-cell NF-κB oscillations synchronize to an oscillating TNF signal in a wide range of stimulation frequencies and become entrained. Entrainment causes amplification of NF-κB oscillations and increased gene expression (Figure 7). Simulations predict that both intrinsic and extrinsic noise can improve NF-κB entrainment range, allowing cells to respond synchronously to broader range of inflammatory signals (Figure 4). Single-cell measurements confirmed that indeed NF-κB entrainment occurs in the broad range as predicted by stochastic modeling (Figure 5). While extrinsic noise leads to differences in oscillation frequency between cells

creating heterogeneous locking behavior and increases entrainment robustness of the population to changes in input period, a surprising finding is the beneficial role for intrinsic noise in dynamical signaling: molecular fluctuation arising from low copy-number feedback transcripts (IκB and A20) can act to enhance NF-κB oscillation and expand the range of inputs that entrain NF-κB and ultimately enhance target gene expression (Figure 7). Increased gene expression was explained by incorporating data on non-linear NF-κB-DNA binding affinity into the model, and we see the greatest differential regulation for late genes such as Ccl5 in agreement with findings that late genes are more sensitive to oscillatory regulation (Ashall et al., 2009; Wee et al., 2012).

Together, our results describe important functions for oscillation and noise in signaling networks. Transcription factor oscillation allows amplified pathway output in response to a periodic stimulus and thus increases system efficiency by reducing the amount of input signal needed to generate strong response. Oscillation moreover allows control of heterogeneity through

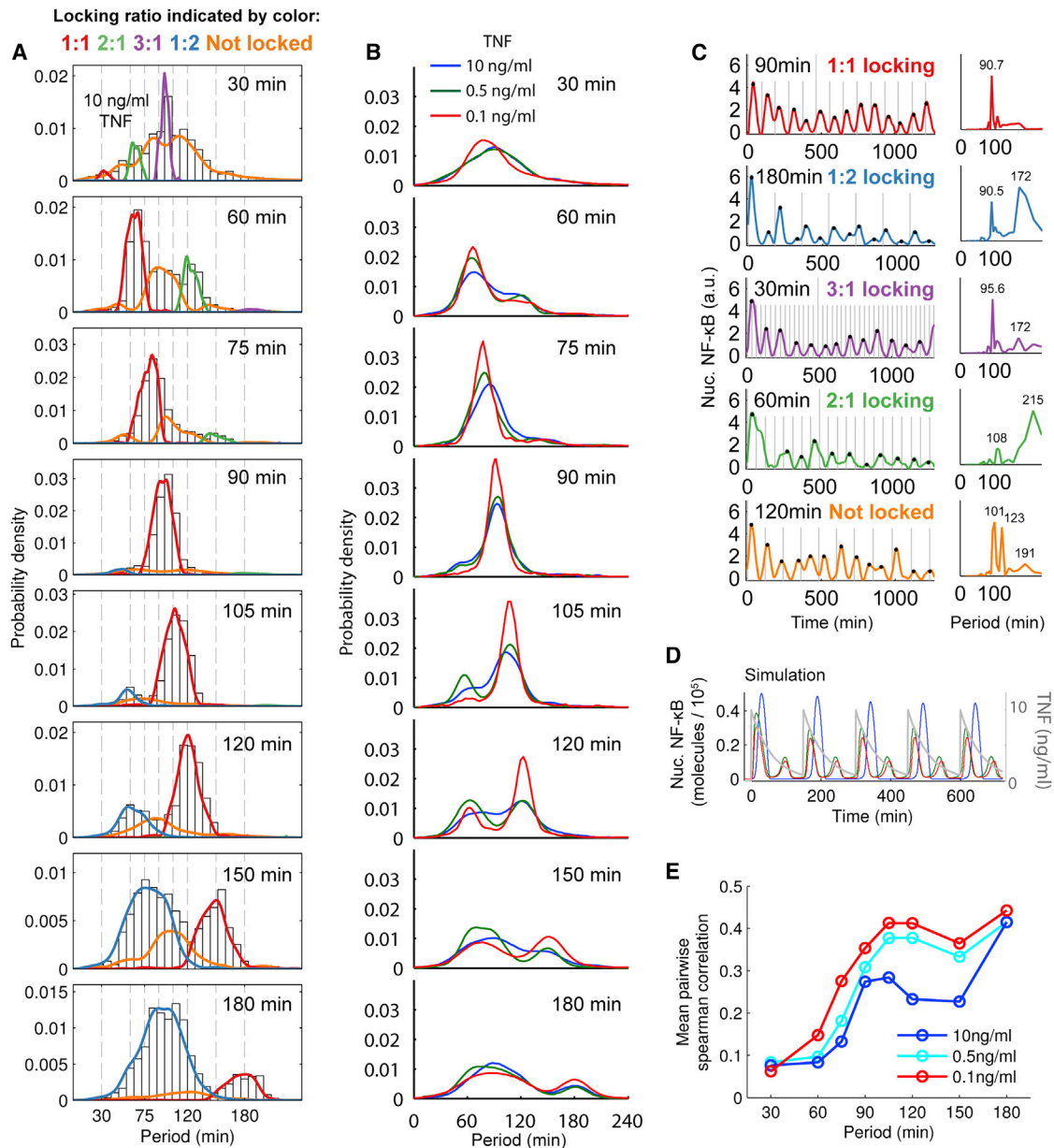


Figure 6. Population Heterogeneity and Dose Dependence of NF-κB Entrainment

(A) Period probability distributions for 10 ng/ml TNF input reveal entrainment at multiple ratios between the input and output period. Under 90 min input, the population entrains nearly homogeneously with a 90 min phase-locked oscillation (1:1 ratio, red line). In contrast, during 150 min stimulation cells may respond with a 150 min oscillation (1:1 ratio, red line), or a 75 min oscillation (1:2 ratio, blue line), or without phase-locking (orange line).

(B) Period distributions for multiple TNF concentrations. Lower concentration leads to period distribution narrowing, indicating improved entrainment and reduced cell-to-cell variability.

(C) Measured single-cell NF-κB traces and power spectra for each locking ratio and an example quasiperiodic response (Not locked).

(D) Simulation with extrinsic noise shows that different locking ratios may occur simultaneously (blue line – 1:1 locking, red and green lines – 1:2 locking).

(E) Mean pairwise spearman correlation in NF-κB dynamics indicating better population entrainment at lower input concentration and at higher input periods. The NF-κB system efficiently filters rapid input fluctuations with periods shorter than 50 min.

See also Figure S6.

synchronization of gene regulatory dynamics across the population. By enhancing oscillation and entrainment bandwidth, noise facilitates efficient transcription in dynamic signaling contexts.

Cytokines like TNF activate multiple signaling pathways, and resonant pathway stimulation provides a way to achieve specific responses. A low-dose signal, delivered periodically, could excite NF-κB oscillations and activate NF-κB signaling while

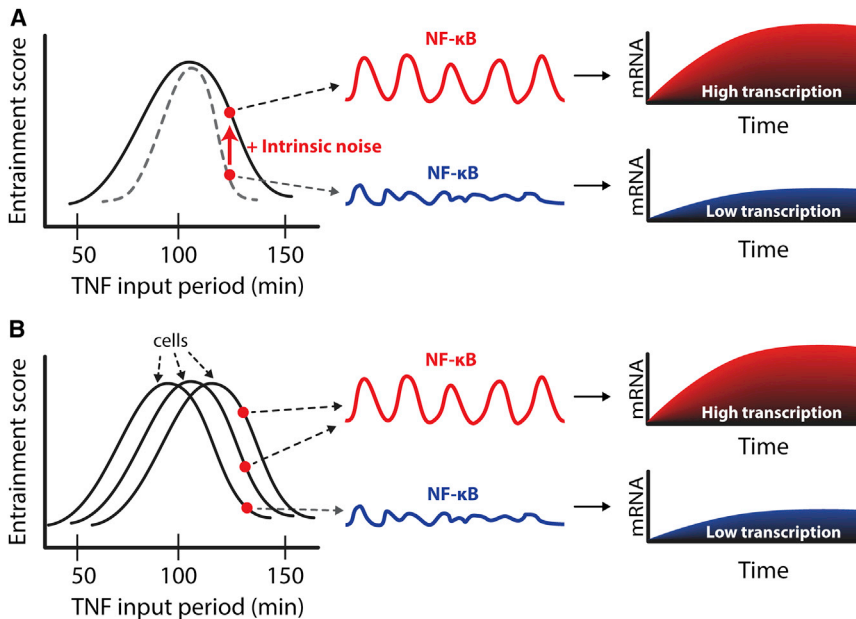


Figure 7. Role of Intrinsic and Extrinsic Noise in NF- κ B Entrainment and Enhanced Gene Expression

(A and B) Entrainment score for different inputs shown on the left side; single-cell NF- κ B time course shown in the middle; and the corresponding mRNA output is shown on the right. (A) Black curve on the left shows the entrainment range of a given cell with intrinsic noise, and dashed gray curve shows the narrower noise-free entrainment range. Signaling inputs at the edge of the entrainment range cause non-entrained NF- κ B responses and small amplitude (in blue), resulting in impaired gene expression output. Intrinsic noise improves the amplitude and the regularity of NF- κ B oscillations (in red), resulting in increased gene expression output. Intrinsic noise can increase entrainment score and also the bandwidth, where cells entrain to a broader range of input periods. (B) Extrinsic noise creates cell-to-cell variability in the entrainment range, resulting in a broader entrainment bandwidth for the population. Population variability in entrainment potential ensures that at least some cells will entrain under a given input period.

avoiding activation of non-oscillatory pathways (such as AP-1). Entrainment with resonance also allows more efficient communication. Indeed, we show that a periodic resonant stimulus achieves greater pathway output while at the same time requiring fewer TNF molecules than a non-entraining stimulus.

Oscillation with resonance may act as a filter. Non-entraining inputs like rapid TNF fluctuations are effectively attenuated at the gene expression level. This may allow NF- κ B system to filter out fast cytokine fluctuations that are not physiological (i.e., input noise). While NF- κ B exhibits a robust natural period of ~ 90 min, researchers are finding oscillation in many signaling pathways with differing characteristic frequencies. Therefore the pathway specificity of a pleiotropic factor such as TNF might be tuned by changing the frequency with which it stimulates a cell. Nonetheless, it is likely that oscillatory pathways are linked within and between cells more than is currently appreciated (Kupzig et al., 2005), and temporal filtering allows cells to achieve specific responses based on the frequency content of input signals.

NF- κ B both responds to and drives cytokine production, and oscillatory cytokine production has been observed in activated single T cells (Han et al., 2012). Entrainment of NF- κ B could be a coordination mechanism during infection, by controlling paracrine signals that instruct migration or fate determination of immune cells (Yde et al., 2011). TNF-positive feedback in secretory immune cells such as macrophages could improve entrainment at higher cell density, creating a more amplified (and more homogeneous) response (Pekalski et al., 2013). The broad entrainment range of the NF- κ B system allows cells to adapt their oscillation frequency and gene expression dynamics to match cytokine fluctuation in the environment.

Our findings suggest a surprising role for noise and oscillation in mammalian signal transduction and transcriptional control (Figure 7). In dynamic, physiological signaling scenarios oscillation provides cells the ability to decode not only the amplitude

but also frequency content of input signals. Inputs occurring near the natural frequency of an oscillatory system are amplified and generate higher gene expression output, while other input frequencies generate an attenuated response. By enhancing oscillation and entrainment at small signal modulation noise may improve the transfer of weaker dynamic signals in the NF- κ B system. Entrainment allows efficient cell-cell communication, control of cell-cell heterogeneity, and possibility to selectively activate oscillatory pathways through resonant stimulation. The prevalence of oscillation in signaling networks suggests that cells are well-equipped for processing dynamic signals.

EXPERIMENTAL PROCEDURES

TNF- α Stimulation Using Microfluidic Cell Culture

We use the cell culture chip described previously (Gómez-Sjöberg et al., 2007). Cells were seeded in PDMS chambers coated with fibronectin at constant density $\sim 20,000$ cells/cm² and were cultured overnight prior to stimulation. Standard culture conditions of 5% CO₂ and 37°C were maintained using an incubation chamber. Mouse TNF- α (Invitrogen) was diluted in DMEM media in vials pressured with 5% CO₂ and kept on ice. Microbore tubing (PEEK, IDEX) connected the TNF- α supply to the chip. For continuous pumping input, the on-chip peristaltic pump was operated at a flow rate ~ 200 nl/min. For periodic input, TNF- α containing media was introduced and incubated in the chamber, allowing degradation and internalization of the ligand. The chamber volume is replaced with fresh TNF- α containing media at defined intervals, leading to periodic sawtooth pattern in ligand concentration.

Cell Retrieval and Gene Expression Analysis

Cells were loaded into the cell culture chip, and a Matlab program delivered 60 min or 120 min periodic inputs with start times staggered by 30 min to generate time points from 0 to 23.5 hr. Cells in one chamber (approximately 200 cells) were retrieved for each time point. At the conclusion of stimulation, cells in all chambers were lysed at once on-chip, and retrieved in a 2 μ l volume of lysis buffer using an automated routine. Cells exited the chip through ~ 10 cm length microbore tubing positioned into wells of a 96-well plate. Wash steps using PBS prior to retrieval prevented cross-contamination of

chambers. cDNA was synthesized using Cells Direct One Step RT-PCR kit (Invitrogen). TaqMan primers and probes (Applied Biosystems) were used for real-time qPCR. Gene expression was assayed using the 48.48 Dynamic Array IFC chip (Fluidigm). Cycle thresholds (CT) were converted to relative expression values normalized to GAPDH ($2^{CT_{gapdh} - CT_{gene}}$). Total expression abundance was calculated as the integral of the relative expression (using the Matlab *trapz* function).

Image Acquisition and Data Processing

The microfluidic chip was mounted on an automated Leica DMI6000B microscope, and fluorescence images (red and green channels for p65 and H2B reporters, respectively) were acquired at 20× magnification via a Retiga-SRV CCD camera (QImaging) every 5–6 min for 24–48 hr. CellProfiler software (www.cellprofiler.org) and custom Matlab routines (Gómez-Sjöberg et al., 2007) were used for image processing (available on request). NF-κB activation was quantified as mean nuclear fluorescence intensity after background correction. Area-under-curve provides a measure of total NF-κB activity (Tay et al., 2010) and was quantified as the integral of the NF-κB response (using Matlab *trapz*). For peak analysis and heatmaps data were smoothed and standardized (Matlab functions *smooth* and *zscore*) followed by peak detection (Matlab *mspeaks*). Peak-to-peak distances were computed as the difference between peak times (Matlab *diff*). Cell image overlays aided visualization of oscillation peaks (colored green) and troughs (colored red).

NF-κB Reporter Cell Line

Mouse (3T3) fibroblasts expressing near-endogenous p65 levels were described previously (Tay et al., 2010). Briefly, p65^{-/-} mouse 3T3 fibroblasts were engineered to express p65-DsRed under control of 1.5 kb p65 promoter sequence (Lee et al., 2009; Tay et al., 2010). A clone was selected with minimum detectable fluorescence intensity to achieve near-endogenous expression level and NF-κB dynamics similar to wild-type (Lee et al., 2009). Addition of ubiquitin-promoter driven H2B-GFP expression provided a nuclear label to facilitate automated tracking and image processing.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and four movies and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.01.013>.

ACKNOWLEDGMENTS

We acknowledge Tomasz Lipniacki, Mustafa Khammash, Benjamin Hepp, and Ankit Gupta for discussions on NF-κB and transcription modeling, and Kobi Benenson and Alexander Hoffmann for commenting on the manuscript. S.T. acknowledges European Union ERC Starting Grant (SingleCellDynamics), Swiss Nationalfonds, SNF SystemsX, and NCCR Molecular Systems Engineering for funding support.

Received: August 18, 2014
Revised: November 2, 2014
Accepted: January 5, 2015
Published: January 29, 2015

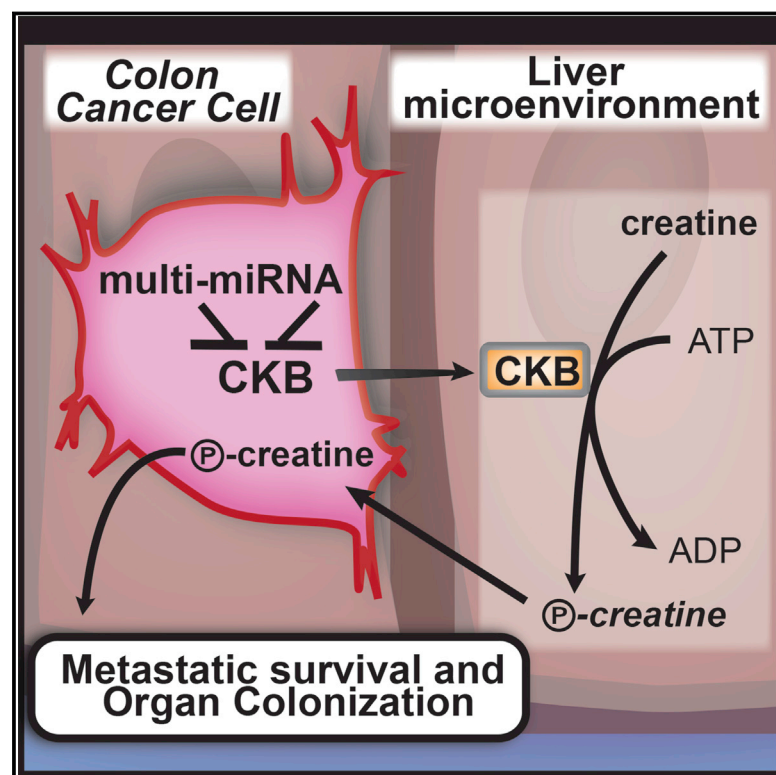
REFERENCES

- Abraham, U., Granada, A.E., Westermark, P.O., Heine, M., Kramer, A., and Herzog, H. (2010). Coupling governs entrainment range of circadian clocks. *Mol. Syst. Biol.* 6, 438.
- Albeck, J.G., Mills, G.B., and Brugge, J.S. (2013). Frequency-modulated pulses of ERK activity transmit quantitative proliferation signals. *Mol. Cell* 49, 249–261.
- Ashall, L., Horton, C.A., Nelson, D.E., Paszek, P., Harper, C.V., Sillitoe, K., Ryan, S., Spiller, D.G., Unitt, J.F., Broomhead, D.S., et al. (2009). Pulsatile stimulation determines timing and specificity of NF-κB-dependent transcription. *Science* 324, 242–246.
- Berridge, M.J., Bootman, M.D., and Roderick, H.L. (2003). Calcium signalling: dynamics, homeostasis and remodelling. *Nat. Rev. Mol. Cell Biol.* 4, 517–529.
- Cai, L., Dalal, C.K., and Elowitz, M.B. (2008). Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* 455, 485–490.
- Cheong, R., Rhee, A., Wang, C.J., Nemenman, I., and Levchenko, A. (2011). Information transduction capacity of noisy biochemical signaling networks. *Science* 334, 354–358.
- Collins, J.J., Imhoff, T.T., and Grigg, P. (1996). Noise-enhanced information transmission in rat SA1 cutaneous mechanoreceptors via aperiodic stochastic resonance. *J. Neurophysiol.* 76, 642–645.
- Dolmetsch, R.E., Xu, K., and Lewis, R.S. (1998). Calcium oscillations increase the efficiency and specificity of gene expression. *Nature* 392, 933–936.
- Douglass, J.K., Wilkens, L., Pantazeli, E., and Moss, F. (1993). Noise enhancement of information transfer in crayfish mechanoreceptors by stochastic resonance. *Nature* 365, 337–340.
- Eldar, A., and Elowitz, M.B. (2010). Functional roles for noise in genetic circuits. *Nature* 467, 167–173.
- Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. *Science* 297, 1183–1186.
- Erzberger, A., Hampp, G., Granada, A.E., Albrecht, U., and Herzog, H. (2013). Genetic redundancy strengthens the circadian clock leading to a narrow entrainment range. *J. R. Soc. Interface* 10, 20130221.
- Falcke, M. (2003). Deterministic and stochastic models of intracellular Ca²⁺ waves. *New J. Phys.* 5, 96.
- Gérard, C., and Goldbeter, A. (2012). Entrainment of the mammalian cell cycle by the circadian clock: modeling two coupled cellular rhythms. *PLoS Comput. Biol.* 8, e1002516.
- Goldbeter, A., and Pourquie, O. (2008). Modeling the segmentation clock as a network of coupled oscillations in the Notch, Wnt and FGF signaling pathways. *J. Theor. Biol.* 252, 574–585.
- Goldbeter, A., Li, Y., and Dupont, G. (1990). Oscillatory dynamics in intercellular communication. *Biomed. Biochim. Acta* 49, 935–940.
- Gómez-Sjöberg, R., Leyrat, A.A., Pirone, D.M., Chen, C.S., and Quake, S.R. (2007). Versatile, fully automated, microfluidic cell culture system. *Anal. Chem.* 79, 8557–8563.
- Han, Q., Bagheri, N., Bradshaw, E.M., Hafner, D.A., Lauffenburger, D.A., and Love, J.C. (2012). Polyfunctional responses by human T cells result from sequential release of cytokines. *Proc. Natl. Acad. Sci. USA* 109, 1607–1612.
- Hao, S., and Baltimore, D. (2013). RNA splicing regulates the temporal order of TNF-induced gene expression. *Proc. Natl. Acad. Sci. USA* 110, 11934–11939.
- Hayden, M.S., and Ghosh, S. (2008). Shared principles in NF-κB signaling. *Cell* 132, 344–362.
- Hoffmann, A., Levchenko, A., Scott, M.L., and Baltimore, D. (2002). The IκB-α-NF-κB Signaling Module: Temporal Control and Selective Gene Activation (New York, NY: Science), pp. 1–5.
- Huh, D., and Paulsson, J. (2011). Random partitioning of molecules at cell division. *Proc. Natl. Acad. Sci. USA* 108, 15004–15009.
- Jensen, M.H., and Krishna, S. (2012). Inducing phase-locking and chaos in cellular oscillators by modulating the driving stimuli. *FEBS Lett.* 586, 1664–1668.
- Junkin, M., and Tay, S. (2014). Microfluidic single-cell analysis for systems immunology. *Lab Chip* 14, 1246–1260.
- Kageyama, R., Ohtsuka, T., Shimojo, H., and Imayoshi, I. (2008). Dynamic Notch signaling in neural progenitor cells and a revised view of lateral inhibition. *Nat. Neurosci.* 11, 1247–1251.
- Kellogg, R.A., Gómez-Sjöberg, R., Leyrat, A.A., and Tay, S. (2014). High-throughput microfluidic single-cell analysis pipeline for studies of signaling dynamics. *Nat. Protoc.* 9, 1713–1726.
- Kupzig, S., Walker, S.A., and Cullen, P.J. (2005). The frequencies of calcium oscillations are optimized for efficient calcium-mediated activation of Ras and the ERK/MAPK cascade. *Proc. Natl. Acad. Sci. USA* 102, 7577–7582.

- Lee, T.K., Denny, E.M., Sanghvi, J.C., Gaston, J.E., Maynard, N.D., Hughey, J.J., and Covert, M.W. (2009). A noisy paracrine signal determines the cellular NF-kappaB response to lipopolysaccharide. *Sci. Signal.* 2, ra65.
- Leloup, J.C., and Goldbeter, A. (2003). Toward a detailed computational model for the mammalian circadian clock. *Proc. Natl. Acad. Sci. USA* 100, 7051–7056.
- Levine, J.H., Lin, Y., and Elowitz, M.B. (2013). Functional roles of pulsing in genetic circuits. *Science* 342, 1193–1200.
- Lindner, B., Garcia-Ojalvo, J., Neiman, A., and Schimansky-Geier, L. (2004). Effects of noise in excitable systems. *Phys. Rep.* 392, 321–424.
- Lipniacki, T., Paszek, P., Brasier, A.R., Luxon, B., and Kimmel, M. (2004). Mathematical model of NF-kappaB regulatory module. *J. Theor. Biol.* 228, 195–215.
- Lipniacki, T., Puszynski, K., Paszek, P., Brasier, A.R., and Kimmel, M. (2007). Single TNFalpha trimers mediating NF-kappaB activation: stochastic robustness of NF-kappaB signaling. *BMC Bioinformatics* 8, 376.
- Little, J.W., Shepley, D.P., and Wert, D.W. (1999). Robustness of a gene regulatory circuit. *EMBO J.* 18, 4299–4307.
- Longo, D.M., Selimkhanov, J., Kearns, J.D., Hasty, J., Hoffmann, A., and Tsimring, L.S. (2013). Dual delayed feedback provides sensitivity and robustness to the NF-kB signaling module. *PLoS Comput. Biol.* 9, e1003112.
- McDonnell, M.D., and Ward, L.M. (2011). The benefits of noise in neural systems: bridging theory and experiment. *Nat. Rev. Neurosci.* 12, 415–426.
- Mondragón-Palomino, O., Danino, T., Selimkhanov, J., Tsimring, L., and Hasty, J. (2011). Entrainment of a population of synthetic genetic oscillators. *Science* 333, 1315–1319.
- Mori, T., and Kai, S. (2002). Noise-induced entrainment and stochastic resonance in human brain waves. *Phys. Rev. Lett.* 88, 218101.
- Nelson, D.E., Ihekweaba, A.E.C., Elliott, M., Johnson, J.R., Gibney, C.A., Foreman, B.E., Nelson, G., See, V., Horton, C.A., Spiller, D.G., et al. (2004). Oscillations in NF-kappaB signaling control the dynamics of gene expression. *Science* 306, 704–708.
- Paszek, P., Ryan, S., Ashall, L., Sillitoe, K., Harper, C.V., Spiller, D.G., Rand, D.A., and White, M.R. (2010). Population robustness arising from cellular heterogeneity. *Proc. Natl. Acad. Sci. USA* 107, 11644–11649.
- Pękalski, J., Zuk, P.J., Kocharczyk, M., Junkin, M., Kellogg, R., Tay, S., and Lipniacki, T. (2013). Spontaneous NF-kB activation by autocrine TNF α signaling: a computational analysis. *PLoS ONE* 8, e78887.
- Perc, M., and Marhl, M. (2003). Noise enhances robustness of intracellular Ca²⁺ oscillations. *Phys. Lett. A* 316, 304–310.
- Phelps, C.B., Sengchanthalangsy, L.L., Malek, S., and Ghosh, G. (2000). Mechanism of kappa B DNA binding by Rel/NF-kappa B dimers. *J. Biol. Chem.* 275, 24392–24399.
- Pikovsky, A., Rosenblum, M., and Kurths, J. (2003). Synchronization: a universal concept in nonlinear sciences, *Volume 12* (Cambridge: Cambridge University Press).
- Purvis, J.E., and Lahav, G. (2013). Encoding and decoding cellular information through signaling dynamics. *Cell* 152, 945–956.
- Purvis, J.E., Karhohs, K.W., Mock, C., Batchelor, E., Loewer, A., and Lahav, G. (2012). p53 dynamics control cell fate. *Science* 336, 1440–1444.
- Reppert, S.M., and Weaver, D.R. (2002). Coordination of circadian timing in mammals. *Nature* 418, 935–941.
- Schütze, J., Mair, T., Hauser, M.J.B., Falcke, M., and Wolf, J. (2011). Metabolic synchronization by traveling waves in yeast cell layers. *Biophys. J.* 100, 809–813.
- Shankaran, H., Ippolito, D.L., Chrisler, W.B., Resat, H., Bollinger, N., Opresko, L.K., and Wiley, H.S. (2009). Rapid and sustained nuclear-cytoplasmic ERK oscillations induced by epidermal growth factor. *Mol. Syst. Biol.* 5, 332.
- Süel, G.M., Garcia-Ojalvo, J., Liberman, L.M., and Elowitz, M.B. (2006). An excitable gene regulatory circuit induces transient cellular differentiation. *Nature* 440, 545–550.
- Swain, P.S., Elowitz, M.B., and Siggia, E.D. (2002). Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. USA* 99, 12795–12800.
- Tay, S., Hughey, J.J., Lee, T.K., Lipniacki, T., Quake, S.R., and Covert, M.W. (2010). Single-cell NF-kappaB dynamics reveal digital activation and analogue information processing. *Nature* 466, 267–271.
- Turner, D.A., Paszek, P., Woodcock, D.J., Nelson, D.E., Horton, C.A., Wang, Y., Spiller, D.G., Rand, D.A., White, M.R., and Harper, C.V. (2010). Physiological levels of TNFalpha stimulation induce stochastic dynamics of NF-kappaB responses in single living cells. *J. Cell Sci.* 123, 2834–2843.
- Varela, F., Lachaux, J.P., Rodriguez, E., and Martinerie, J. (2001). The brainweb: phase synchronization and large-scale integration. *Nat. Rev. Neurosci.* 2, 229–239.
- Vilar, J.M., Kueh, H.Y., Barkai, N., and Leibler, S. (2002). Mechanisms of noise-resistance in genetic oscillators. *Proc. Natl. Acad. Sci. USA* 99, 5988–5992.
- Wakamoto, Y., Dhar, N., Chait, R., Schneider, K., Signorino-Gelo, F., Leibler, S., and McKinney, J.D. (2013). Dynamic persistence of antibiotic-stressed mycobacteria. *Science* 339, 91–95.
- Wang, Y., Paszek, P., Horton, C.A., Kell, D.B., White, M.R., Broomhead, D.S., and Muldoon, M.R. (2011). Interactions among oscillatory pathways in NF-kappa B signaling. *BMC Syst. Biol.* 5, 23.
- Wee, K.B., Yio, W.K., Surana, U., and Chiam, K.H. (2012). Transcription factor oscillations induce differential gene expressions. *Biophys. J.* 102, 2413–2423.
- Yang, Q., Pando, B.F., Dong, G., Golden, S.S., and van Oudenaarden, A. (2010). Circadian gating of the cell cycle revealed in single cyanobacterial cells. *Science* 327, 1522–1526.
- Yde, P., Mengel, B., Jensen, M.H., Krishna, S., and Trusina, A. (2011). Modeling the NF-kB mediated inflammatory response predicts cytokine waves in tissue. *BMC Syst. Biol.* 5, 115.
- Yissachar, N., Sharar Fischler, T., Cohen, A.A., Reich-Zeliger, S., Russ, D., Shifrut, E., Porat, Z., and Friedman, N. (2013). Dynamic response diversity of NFAT isoforms in individual living cells. *Mol. Cell* 49, 322–330.
- Yoshiura, S., Ohtsuka, T., Takenaka, Y., Nagahara, H., Yoshikawa, K., and Kagayama, R. (2007). Ultradian oscillations of Stat, Smad, and Hes1 expression in response to serum. *Proc. Natl. Acad. Sci. USA* 104, 11292–11297.
- Zhou, C., Kurths, J., Kiss, I.Z., and Hudson, J.L. (2002). Noise-enhanced phase synchronization of chaotic oscillators. *Phys. Rev. Lett.* 89, 014101.

Extracellular Metabolic Energetics Can Promote Cancer Progression

Graphical Abstract



Authors

Jia Min Loo, Alexis Scherl, ..., Philip B. Paty, Sohail F. Tavazoie

Correspondence

stavazoie@mail.rockefeller.edu

In Brief

The tumor microenvironment is a rich source of metabolic substrates that could be utilized by cancer cells. Colon cancer cells secrete a kinase that acts extracellularly to generate one such metabolite, phosphocreatinine, that directly fuels tumor growth and metastasis to the liver.

Highlights

- miR-551a and miR-483-5p suppress colon cancer metastasis by targeting CKB
- CKB promote metastatic survival by modulating intra- and extracellular energetics
- miRNAs and CKB can be therapeutically targeted by AAV and small molecule delivery

Accession Numbers

GSE56320



Extracellular Metabolic Energetics Can Promote Cancer Progression

Jia Min Loo,¹ Alexis Scherl,¹ Alexander Nguyen,¹ Fung Ying Man,¹ Ethan Weinberg,¹ Zhaoshi Zeng,² Leonard Saltz,³ Philip B. Paty,² and Sohail F. Tavazoie^{1,*}

¹Laboratory of Systems Cancer Biology, Rockefeller University, New York, NY 10065, USA

²Department of Surgery

³Department of Medicine

Memorial-Sloan Kettering Cancer Center, New York, NY 10065, USA

*Correspondence: stavazoie@mail.rockefeller.edu

<http://dx.doi.org/10.1016/j.cell.2014.12.018>

SUMMARY

Colorectal cancer primarily metastasizes to the liver and globally kills over 600,000 people annually. By functionally screening 661 microRNAs (miRNAs) in parallel during liver colonization, we have identified miR-551a and miR-483 as robust endogenous suppressors of liver colonization and metastasis. These miRNAs convergently target creatine kinase, brain-type (CKB), which phosphorylates the metabolite creatine, to generate phosphocreatine. CKB is released into the extracellular space by metastatic cells encountering hepatic hypoxia and catalyzes production of phosphocreatine, which is imported through the SLC6A8 transporter and used to generate ATP—fueling metastatic survival. Combinatorial therapeutic viral delivery of miR-551a and miR-483-5p through single-dose adeno-associated viral (AAV) delivery significantly suppressed colon cancer metastasis, as did CKB inhibition with a small-molecule inhibitor. Importantly, human liver metastases express higher CKB and SLC6A8 levels and reduced miR-551a/miR-483 levels relative to primary tumors. We identify the extracellular space as an important compartment for malignant energetic catalysis and therapeutic targeting.

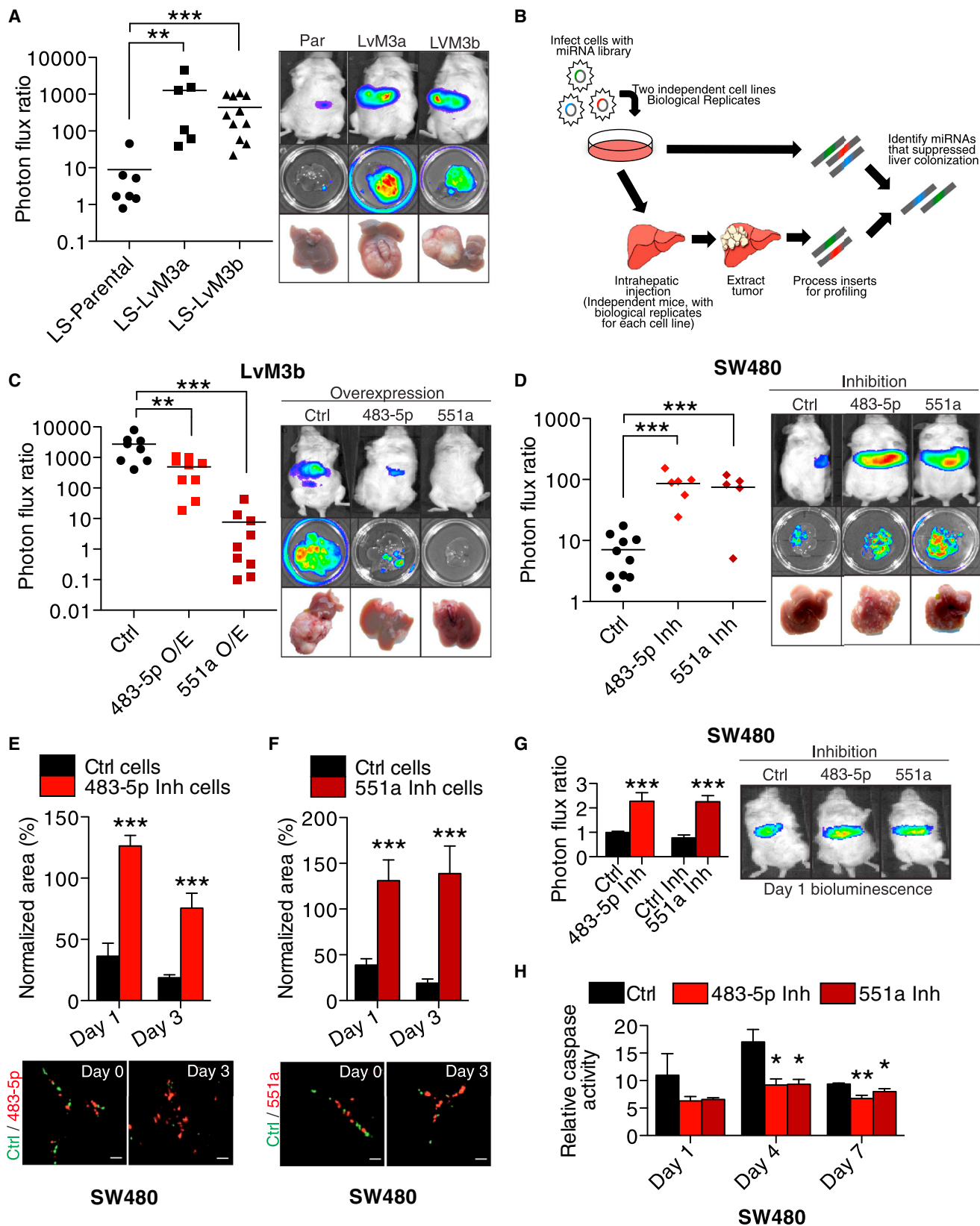
INTRODUCTION

Colorectal cancer is the third leading cause of mortality in the United States and a major cause of death globally (Davis and Schlessinger, 2012; Jemal et al., 2011; Siegel et al., 2014). Death from colorectal cancer is primarily due to the metastatic progression, with the liver being the organ of metastatic colonization in over 70% of patients. To date, efforts aimed at increasing cure rates after surgery have focused on combination chemotherapy administration as a means of preventing metastasis. Such therapy reduces metastatic relapse by roughly 7% (Meyerhardt and Mayer, 2005). The high prevalence of this disease and the lack of effective adjuvant therapeutics demand a greater understanding

of the biology of its progression (Markowitz and Bertagnoli, 2009).

In recent years, posttranscriptional deregulation has emerged as a key feature of metastatic cells. In particular, specific microRNAs (miRNAs), which are small noncoding RNAs, have been identified that are silenced or overexpressed and act to suppress or promote metastatic progression by diverse cancer types (Lujambio and Lowe, 2012; Ma et al., 2007; Pencheva and Tavazoie, 2013; Pencheva et al., 2012; Tavazoie et al., 2008). While the use of these miRNAs as molecular probes for the identification of metastasis regulators has proved fruitful, their therapeutic utility has been limited given the inefficient delivery of miRNAs into various metastatic tissues. Interestingly, the liver represents an exception to this rule, because miRNAs tend to accumulate in hepatic tissue and because vectors such as adeno-associated viruses and nanoparticles have shown promising efficacy in enhancing hepatic delivery in nonhuman primates and humans (Kota et al., 2009; Mingozi and High, 2011). Given this unique feature of the liver as well as the great need for targeted therapies that can suppress liver metastatic colonization by colon cancer, the identification of miRNAs that could suppress liver metastasis would be of great clinical value.

By screening 661 human miRNAs in parallel for their ability to suppress the colonization of the liver by multiple colon cancer cell lines representing diverse mutational subtypes, we have identified miR-551 and miR-483 as endogenous suppressors of colon cancer metastasis. We find that these miRNAs both target Creatine kinase Brain (CKB). Disseminated metastatic cells release this enzyme into the extracellular space, where it catalyzes the phosphorylation of the metabolite creatine by using extracellular ATP as the phosphate source. Phosphocreatine is then imported into disseminated colorectal cancer cells where its high-energy phosphate is used to generate intracellular ATP that sustains the energetic requirements of colon cancer cells encountering hepatic hypoxia, allowing them to survive this barrier to metastatic progression. Therapeutic viral delivery of these miRNAs to the liver and disseminated colon cancer cells via adeno-associated viral delivery strongly suppresses metastatic colonization by colon cancer cells. Moreover, small-molecule therapeutic inhibition of CKB activity also suppresses metastatic growth. Our findings delineate a druggable molecular network that governs both the metabolic state and the metastatic progression capacity of disseminated colon cancer cells. More



(legend on next page)

importantly, we implicate the extracellular space as a previously unrecognized environment for malignant catalysis and identify CKB as a secreted metabolic kinase that drives cancer progression.

RESULTS

Endogenous miR-483-5p and miR-551a Suppress Human Colorectal Cancer Metastasis

In vivo selection has been used by many investigators to identify candidate genes that regulate metastatic progression of diverse cancer types. This approach allows one to derive highly metastatic subpopulations with enhanced metastatic activity for a given organ (Fidler, 1973). The comparison of transcriptomic profiles of metastatic derivatives to the parental lines from which they were derived has revealed numerous candidate genes for functional testing (Brunns et al., 1999; Kang et al., 2003; Minn et al., 2005; Pencheva et al., 2012; Png et al., 2012; Tavazoie et al., 2008). As a first step to identify the molecular regulators of liver colonization by colon cancer cells, we performed in vivo selection on the LS-174T (K-Ras mutant) human colon cancer line for enhanced liver colonization activity through iterative intra-hepatic injection of cancer cells into immunodeficient mice followed by surgical resection of liver colonies and dissociation of cells. Independently derived third-generation liver colonizers LS-LvM3a and LS-LvM3b displayed significantly enhanced (>50-fold) capacity for liver colonization upon intrahepatic injection relative to their parental line (Figure 1A). Importantly, these derivatives also displayed dramatically enhanced (>150-fold) liver metastatic capacity upon portal circulation injection in metastasis assays (Figure S1A available online)—revealing the acquisition of liver colonization capacity to be sufficient for imparting enhanced liver metastasis activity. As an orthogonal approach, we transduced a library of lentiviral particles, each encoding one of 661 human miRNAs, into two independent colon cancer cell lines—the WiDr (K-Ras wild-type) and SW620 (K-Ras mutant) human lines. These cancer populations, containing cancer cells expressing each of 661 miRNAs, were then intrahepatically injected into mice to allow for selection of cells capable of colonizing the liver. Genomic PCR amplification of lentiviral-derived miRNA sequences and miRNA profiling of miRNA inserts

allowed for the quantification of miRNA insert representation (Figure 1B; Table S1). We identified miRNAs that displayed reduced representation in the context of liver colonization in both colon cancer cell-lines on the basis that overexpression of these miRNAs suppressed liver colonization by colon cancer cells. We next asked whether endogenous forms of these miRNAs exhibited silencing in highly metastatic derivatives relative to isogenic poorly metastatic parental cells. Indeed, two of the miRNAs, miR-483-5p and miR-551a, were found to be silenced in highly metastatic LS-LvM3a and LS-LvM3b liver colonizers relative to their parental line (Figure S1B; Table S2). Consistent with a suppressive role for these miRNAs in liver colonization, overexpression of miR-483-5p or miR-551a robustly suppressed metastatic colonization by LS-LvM3b cells introduced into the portal circulation (Figures 1C and S1C), while inhibition of endogenous miR-483-5p or miR-551a in poorly metastatic parental lines SW480 and LS-174T significantly enhanced liver metastatic colonization (Figures 1D and S1D). The effects of these miRNAs on metastatic progression were not secondary to modulation of proliferative capacity because miR-551a inhibition did not affect in vitro proliferation, while miR-483-5p inhibition minimally increased proliferation (10%)—an order of magnitude less than its effect on metastasis (Figure S1E). Importantly, overexpression of either miRNA did not suppress primary tumor growth (Figure S1F).

To better investigate the mechanism(s) by which these miRNAs exert their anti-metastatic effects, we employed an in vitro liver organotypic slice culture system to study early events during liver metastasis subsequent to single-cell dissemination of colon cancer cells in the liver microenvironment (Figure S1G). Consistent with prior studies, which revealed a significant selection on cell survival during metastatic colonization (Gupta and Massagué, 2006; Talmadge and Fidler, 2010), we noted that highly metastatic LvM3b cells were significantly better at persisting in the liver microenvironment than their poorly metastatic parental line; consistent with a key role for intrahepatic persistence in metastatic progression (Figure S1H). We examined whether the enhanced capacity of metastatic cells to persist in the hepatic microenvironment is regulated by miR-483-5p or miR-551a. Indeed, overexpression of miR-483-5p and miR-551a in LS-LvM3b cells suppressed (Figures S1I and S1J), while

Figure 1. miR-483-5p and miR-551a Are Endogenous miRNAs that Suppress Liver Metastasis

- (A) Bioluminescence plot of liver colonization by 5×10^5 LS-Parental, LvM3a, and LvM3b cells after direct intrahepatic injection ($n > 5$). Mice were imaged at day 21 after injection and livers extracted for ex vivo imaging and gross morphological examination. Photon flux ratio is the ratio of bioluminescence signal at day 21 normalized to signal on day 0.
- (B) Schematic for the identification of miR-483-5p and miR-551a as suppressors of metastasis.
- (C) Liver metastasis of mice injected with 5×10^5 LvM3b cells overexpressing either a control hairpin, miR-483-5p, or miR-551a ($n > 5$).
- (D) Liver metastasis in mice injected with 5×10^5 SW480 cells, whose endogenous miR-483-5p or miR-551a was inhibited ($n > 5$).
- (E and F) Organotypic slice culture imaging of SW480 cells ($n = 8$) whose endogenous miR-483-5p (E) or miR-551a (F) were inhibited by LNAs. Cells (5×10^5) were labeled with cell-tracker green (control LNA) or cell-tracker red (miRNA-specific LNA) and introduced into the livers prior to slice culture. Dye-swap experiments were performed to compensate for dye bias. Representative images at day 0 and day 3 are shown. Total area of each cell population at indicated time points are measured and normalized to start of experiment. Scale bar represents 50 μm .
- (G) Bioluminescent metastatic signal from mice ($n = 5$) injected with 5×10^5 SW480 cells whose endogenous miR-483-5p or miR-551a activities were inhibited. Images and measurements were taken 24 hr after tumor cells inoculation.
- (H) Relative in vivo caspase activity of SW480 cells whose endogenous miR-483-5p or miR-551a was inhibited ($n = 3$). Caspase activity was monitored using a caspase-3/7-activated DEVD-luciferin and normalized with bioluminescent signal from regular luciferin. Error bars represent SEM; all p values are based on one-sided Student's t tests, or where appropriate, Mann-Whitney test for non-Gaussian distribution. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.
- See also Figure S1 and Tables S1 and S2.

inhibition of these miRNAs enhanced (Figures 1E and 1F) colon cancer persistence in the hepatic microenvironment. In agreement with our organotypic slice culture findings, we noted that as early as 24 hr after injection of cells into the portal circulation, cells whose endogenous miRNAs were inhibited out-competed control cells (Figure 1G). As neither of these miRNAs significantly affected proliferation, we asked if they elicited their effects by suppressing cancer cell survival during metastatic progression. To quantify cell death *in vivo*, we utilized a bioluminescence-based luciferin reporter of caspase-3/7 activity (Hickson et al., 2010). MiRNA inhibition significantly reduced *in vivo* caspase activity in colon cancer cells during the early phase of hepatic colonization (Figure 1H), revealing cancer survival to be the phenotype suppressed by these miRNAs. These *in vivo* findings provide corroboration and a mechanistic basis for the organotypic slice culture observations. Our findings reveal miR-483 and miR-551a to suppress liver metastatic colonization through suppression of metastatic cell survival in the liver—a phenotype exhibited by highly metastatic colon cancer cells.

miR-483-5p and miR-551a Suppress Colorectal Cancer Cell Survival and Metastasis in the Liver through Targeting of CKB

We next sought to identify the downstream effectors of these miRNAs. Through transcriptomic profiling, we identified transcripts that were downregulated by overexpression of each miRNA and that contained 3'-UTR or coding-sequence elements complementary to the miRNAs. Interestingly, CKB was identified as a putative target of both miRNAs, suggesting that these miRNAs, which exhibit common *in vivo* and organotypic phenotypes, might mediate their effects through a common target gene (Table S3). Importantly, endogenous miR-483 and miR-551a were found to suppress CKB protein levels (Figure 2A). Quantitative RT-PCR validation also revealed suppression or upregulation of CKB transcript levels upon overexpression or inhibition of the miRNAs, respectively (Figures S2A and S2B). Mutagenesis and luciferase-based reporter assays revealed miR-483 to directly target the 3' UTR and miR-551a to directly target the coding region of CKB (Figures S2C and S2D). Overexpression of CKB in poorly metastatic SW480 cells was sufficient to significantly enhance liver metastasis (>3-fold; Figure 2B), while CKB knockdown in metastatic LS-LvM3b cells and SW480 cells, through the use of two independent hairpins for each line, robustly suppressed liver metastatic colonization (>5-fold; Figures 2C and S2E). Importantly, metastases that grew out in knockdown experiments had “escaped” small hairpin RNA (shRNA) knockdown and displayed restored CKB expression (Figure S2F). Consistent with the effects of its regulatory miRNAs, CKB overexpression was sufficient to significantly enhance the ability of colon cancer cells to persist in the liver microenvironment and enhanced their representation in the liver (Figure 2D), while CKB knockdown substantially reduced intrahepatic persistence (Figure 2E). Consistent with this, CKB overexpression reduced (Figure 2F), while CKB knockdown significantly enhanced (Figure 2G), *in vivo* caspase-3/7 activity in colon cancer cells during the initial phase of hepatic colonization. To investigate whether CKB acts directly downstream of miR-483-5p and miR-551a, we performed gain-of-function, loss-of-function,

and epistasis studies. Knockdown of CKB in cells inhibited for miR-483-5p or miR-551a prevented the enhanced metastasis effect seen with miR-483-5p or miR-551a silencing (Figure 2H). Conversely overexpression of CKB was sufficient to rescue the suppressed liver metastatic phenotypes of cells overexpressing miR-483-5p or miR-551a (Figure S2G). The results of the above mutational, gain-of-function, loss-of-function, and epistasis analyses reveal CKB to be a direct target of miR-483-5p and miR-551a, to act as a downstream effector of these miRNAs in the regulation of colon cancer metastatic progression, and be a promoter of colon cancer survival during hepatic metastatic colonization.

CKB Promotes Colorectal Cancer Cell Survival during Acute Intrahepatic Hypoxia through Modulation of the Phosphocreatine/ATP Shuttle

CKB is known to regulate the reservoir of rapidly mobilized high-energy phosphates in tissues such as the brain by catalyzing the transfer of a high-energy phosphate group from ATP to the metabolite creatine, yielding phosphocreatine (Wyss and Kaddurah-Daouk, 2000). Recent studies have implicated the involvement of metabolic pathways in tumorigenesis and cancer progression (Cairns et al., 2011; Christofk et al., 2008; Dang et al., 2009; Kaelin and McKnight, 2013). The maintenance of intracellular ATP levels is also critical for cancer cell survival under metabolic stress (Jeon et al., 2012). Highly metabolic cells maintain phosphocreatine stores in order to buffer against low ATP states, because phosphocreatine's high-energy phosphate can be transferred to ADP to generate ATP (Wallimann et al., 1992; Wyss and Kaddurah-Daouk, 2000). Consistent with this, overexpression and knockdown of CKB in colon cancer cells increased and decreased, respectively, intracellular phosphocreatine levels (Figure 3A) and CKB depletion resulted in decreased ATP levels that could be rescued by phosphocreatine supplementation (Figure 3B). Consistent with our findings that miR-483 and miR-551a regulate CKB expression, modulation of either of these miRNAs also modulated intracellular phosphocreatine (Figures S3A and S3B) and ATP levels (Figures S3C and S3D). What purpose could CKB-generated phosphocreatine and ATP play during colon cancer metastatic progression? The liver microenvironment is known to contain hypoxic regions, with metabolically active hepatocytes at the periportal region displaying high rates of oxygen consumption and hepatocytes at the perivenous region actively undergoing glycolysis (Arteel et al., 1995; Jungermann and Kietzmann, 2000). Additionally, colon cancer cells metastasize to the liver via the portal circulation, which is relatively hypoxemic. We hypothesized that colorectal cancer cells experience acute hypoxia and intense competition for glycolytic substrates during initial dissemination to the liver and could be poorly adapted to the liver microenvironment prior to HIF-activated responses (Semenza, 2011). ATP generated from rapid utilization of intracellular phosphocreatine reservoirs might enable colon cancer cells to survive acute hepatic hypoxia. To determine if colon cancer cells experience hypoxia during early metastatic colonization, we utilized a Hif-1 alpha transcriptional luciferase-reporter (HRE-Luc) as an *in vivo* sensor and reporter of cellular hypoxia (Figure S3E) and observed that colon cancer cells experience hypoxia early after hepatic

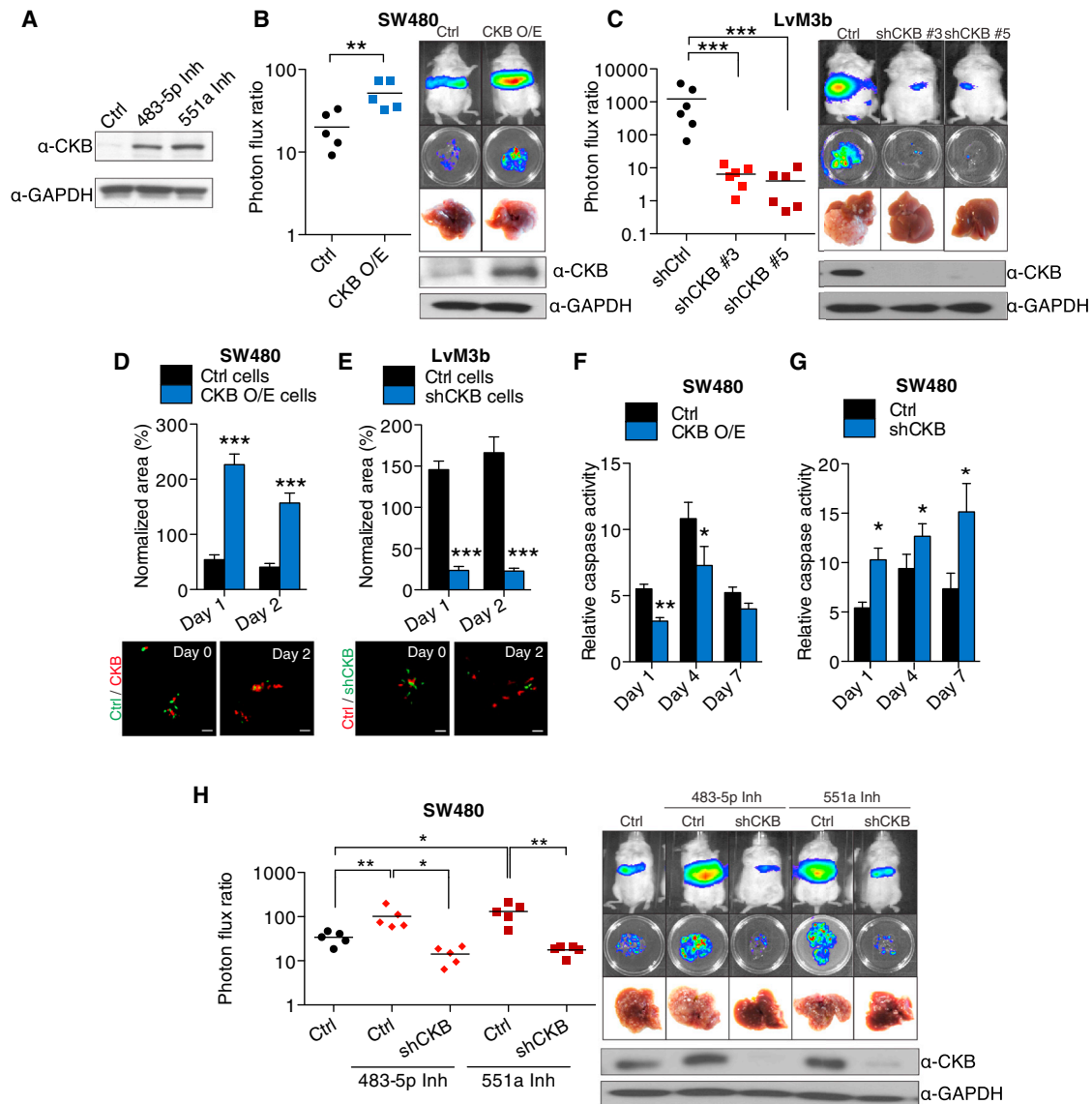


Figure 2. miR-483-5p and miR-551a Suppress Colorectal Cancer Cells Survival and Metastasis in the Liver through Regulation of CKB

(A) Expression of CKB in SW480 cells whose endogenous miR-483-5p or miR-551a was inhibited with LNAs.

(B) Liver metastasis in mice injected intrasplenically with 5×10^5 control SW480 cells or CKB overexpressing SW480 cells ($n = 5$). Mice were euthanized at 28 days after injection.

(C) Liver metastasis in mice injected intrasplenically with 5×10^5 LvM3b expressing a control hairpin or two independent shRNA hairpins targeting CKB ($n = 6$). Mice were euthanized 21 days after injection.

(D) Survival of control SW480 and CKB overexpressing SW480 cells in organotypic liver slices ($n = 8$).

(E) Survival of LvM3b cells expressing a control hairpin or hairpin targeting CKB in organotypic slice cultures ($n = 8$). Representative images at day 0 and day 2 are shown. Scale bar represents 50 μ m.

(F) Relative in vivo caspase activity of control or CKB overexpressing SW480 cells. Caspase activity was measured on days 1, 4, and 7 after injection ($n = 3$).

(G) Relative in vivo caspase-3 activity of SW480 cells expressing a control shRNA or shRNA targeting CKB. Caspase activity was measured on days 1, 4, and 7 after injection ($n = 3$).

(H) Liver metastasis in mice injected with 5×10^5 SW480 cells whose endogenous miR-483-5p or miR-551a was inhibited by LNA, with and without CKB knockdown. Error bars represent SEM; all p values are based on one-sided Student's t tests, or where appropriate, Mann-Whitney test for non-Gaussian distribution. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

See also Figure S2 and Table S4.

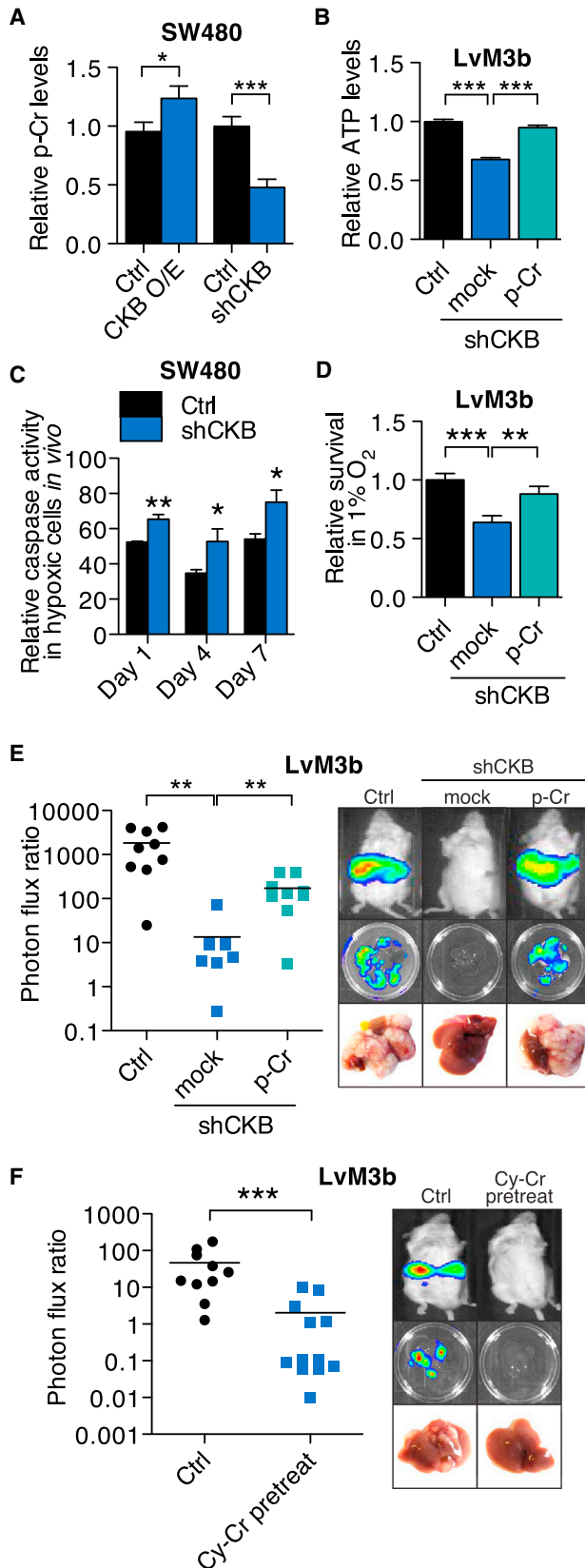


Figure 3. CKB Modulates Colorectal Cancer Cell Survival during Acute Intrahepatic Hypoxia through Modulation of the Phosphocreatine/ATP Shuttle

(A) Relative intracellular phosphocreatine levels in SW480 cells overexpressing CKB or depleted for CKB ($n = 5$). (B) Relative intracellular ATP levels in LvM3b cells depleted for CKB, with and without exogenous 10 μ M phosphocreatine supplementation ($n = 5$). (C) *In vivo* caspase activity of control and CKB knockdown SW480 cells experiencing hypoxia within the livers of mice ($n = 3$). (D) Survival of colorectal cancer cells in hypoxia *in vitro* with and without CKB knockdown and 10 μ M phosphocreatine supplementation ($n = 3$). (E) Liver metastasis by CKB depleted LvM3b cells with and without overnight preincubation with 10 μ M phosphocreatine. Cells (5×10^5) were then inoculated into the liver of mice through intrasplenic injection. (F) Liver metastasis in mice injected with 5×10^5 LvM3b cells with and without pretreatment with 10 mM cyclocreatine for 48 hr. Error bars represent SEM; all p values are based on one-sided Student's t tests, or where appropriate, Mann-Whitney test for non-Gaussian distribution. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

See also Figure S3.

dissemination (Figure S3F). We found that CKB depletion increased caspase-mediated cell death in HRE-Luc expressing cells experiencing hypoxia *in vivo* (Figure 3C). Conversely, inhibition of either miR-483-5p or miR-551a protected HRE-Luc expressing cells experiencing hypoxia *in vivo* (Figure S3G). Consistent with a role for CKB and phosphocreatine in promoting cancer-cell survival during hypoxia, cells depleted of CKB through RNAi displayed reduced survival while experiencing hypoxia *in vitro*—an effect that was abrogated upon phosphocreatine supplementation (Figure 3D). In agreement with our *in vitro* findings, preincubation of colon cancer cells depleted of CKB with phosphocreatine enhanced their ability to metastasize to the liver (Figure 3E). Conversely, liver metastasis was inhibited when we preincubated colon cancer cells with cyclocreatine (Lillie *et al.*, 1993), an inhibitor of CKB that depletes phosphocreatine levels (Figure 3F). Our findings suggest that hepatic hypoxia poses a barrier for colon cancer cells during early metastatic colonization and that cells endure this phase through the generation of ATP from phosphocreatine reserves. Indeed, the ability of phosphocreatine preloading to enhance metastasis *in vivo* supports the importance of the acute initial hypoxic barrier and energetic demands in shaping metastatic colonization by cancer cells as they enter the liver microenvironment.

CKB Is Secreted by Colorectal Cancer Cells and Promotes Malignant Conversion of Extracellular ATP and Creatine to Phosphocreatine

While considering CKB's role in the setting of the hypoxic hepatic microenvironment that colorectal cancer cells arrive into as they disseminate from the gut to the liver, we faced a conundrum: how can colon cancer cells arriving and residing in a hypoxic hepatic microenvironment replenish phosphocreatine if they are depleted of ATP, especially during the acute phase, prior to any hypoxia-response (Bertout *et al.*, 2008; Semenza, 2013; Wheaton and Chandel, 2011)? Earlier clinical studies have described the detection of CKB proteins and CKB activity in the sera of patients with various forms of malignancies and physiological insults (Huddleston *et al.*, 2005; Rubery *et al.*, 1982; Wyss and Kaddurah-Daouk, 2000). The presence of

extracellular ATP in the microenvironment of macrometastases was also reported (Pellegatti et al., 2008; Stagg and Smyth, 2010). Interestingly, the liver is the main synthetic organ for creatine synthesis in the body. We hypothesized that colorectal cancer cells may release CKB into the extracellular space, which can then convert extracellular ATP and liver-produced creatine into phosphocreatine that is then taken up by cancer cells, thereby exerting a protective effect on hypoxic colorectal cancer cells prior to their adaptation to the hypoxic liver microenvironment. We thus investigated the possibility of metastatic colorectal cancer cells releasing CKB extracellularly. Indeed, extracellular CKB was released from metastatic LvM3b cells, but not LvM3b cells expressing a CKB targeting shRNA (Figure 4A). To determine if extracellular CKB was released from live or dying cells, we expressed CKB tagged with a FLAG-epitope through a linker containing a caspase-3/7 recognition DEVD motif. Caspase activation in apoptotic cells would result in caspase recognition and cleavage of the DEVD motif between the FLAG-epitope tag and CKB, causing loss of the FLAG-epitope from the expressed CKB (Figure 4B). We observed that extracellular CKB was released primarily by live cells, as the FLAG-epitope was not lost (Figure 4C). As generation of extracellular phosphocreatine requires exogenous ATP and creatine, we confirmed the presence of extracellular ATP in the microenvironment of incipient micrometastases using a plasma membrane-anchored luciferase reporter for detecting extracellular ATP (pME-Luciferase; Figure 4D) (Pellegatti et al., 2005). If the prometastatic effects of CKB result from utilization of extracellular ATP as a substrate, then depleting extracellular ATP should suppress the prometastatic activity of CKB. Indeed, expressing CD39, a plasma membrane anchored ATP hydrolase in SW480 cells significantly precluded the ability of CKB overexpression to promote metastasis without affecting CKB levels (Figure 4E). Consistent with CKB consumption of extracellular ATP, cells overexpressing CKB or cells whose endogenous miR-483-5p or miR-551a were inhibited displayed significantly lower extracellular ATP levels in vivo relative to the control cells (Figures 4F and S4A). Conversely, the microenvironment surrounding CKB knockdown cells displayed higher extracellular ATP levels (Figure S4B). If extracellular CKB catalysis can enhance metastasis, we reasoned that supplying the product of CKB-mediated catalysis, phosphocreatine, in the extracellular space should rescue the effect of CKB loss-of-function. In order to test this, we implanted a mini-osmotic pump that continuously released phosphocreatine into the peritoneal cavity of immunodeficient mice, which is eventually drained into the portal circulation. Remarkably, exogenous phosphocreatine was sufficient to significantly enhance metastasis (>10-fold) by CKB-depleted cells in vivo (Figure 4G). We next used a Boyden chamber coculture system to determine if colorectal cancer cells overexpressing CKB could promote the survival of CKB knockdown cells under hypoxia (Figure 4H). CKB overexpressing cells were indeed able to compensate for the survival of CKB knockdown cells across the trans-well, while addition of a CKB-activity neutralizing antibody abrogated this effect (Figure 4I). We extended our findings to an in vivo system using colorectal cancer cells depleted of intracellular CKB but expressing a secreted form of CKB fused to the IgK secretory signal sequence. Remarkably,

overexpressing secreted CKB was sufficient to enhance colorectal cancer metastasis (Figure 4J). We further examined the levels of serum CKB in mice injected with CKB knockdown cells. Interestingly, we found that mice with escaped tumors invariably had increased serum CKB levels (Figure S4C).

The Creatine Transporter SLC6a8 Modulates CKB-Mediated Colorectal Cancer Metastasis

Having implicated exogenous creatine/phosphocreatine metabolism in colorectal cancer metastasis, we next investigated the regulation of creatine/phosphocreatine metabolism in colon cancer metastatic progression. Depletion of guanidinoacetate methyltransferase (GAMT), the enzyme required for the final step of creatine synthesis, in colon cancer cells did not affect metastasis (Figure S5A), consistent with a model wherein extracellular (the liver is the primary site of creatine biosynthesis) rather than intracellular creatine drives colorectal cancer metastasis. Next, we asked if SLC6a8, a transporter of creatine compounds (Salomons et al., 2001) modulates phosphocreatine levels in colon cancer cells. Indeed, SLC6a8 knockdown reduced intracellular phosphocreatine and ATP levels (Figure 5A). If extracellular phosphocreatine uptake promotes metastasis, such impairment in phosphocreatine levels should reduce metastasis. Indeed, multiple colon cancer cell lines depleted of SLC6a8 displayed substantially reduced (10- to 100-fold) metastatic activity (Figures 5B and S5B), whereas metastatic tumors that eventually grew out from SLC6a8 knockdown cells were escapers and displayed restored SLC6a8 expression (Figure S5C). Importantly SLC6a8 knockdown, which depleted cellular uptake of extracellular phosphocreatine, also abrogated the effect of CKB overexpression on colorectal cancer metastasis (Figure 5C), revealing extracellular phosphocreatine uptake to be downstream of CKB catalysis. Additionally, depleting SLC6a8 in CKB knockdown cells abrogated the protective effect of phosphocreatine during hypoxic stress (Figure 5D). Finally exogenously added phosphocreatine was not able to promote liver metastasis by SLC6a8 knockdown cells (Figure S5D). These findings reveal SLC6a8 to be downstream of CKB and phosphocreatine and to mediate their metastasis promoting effects.

miR-483-5p, miR-551a, CKB, and SLC6a8 Associate with Human Progression Stage

We next analyzed the expression levels of miR-483 and miR-551a in a set of 66 surgically resected human primary colon cancer and liver metastases from MSKCC. Consistent with a metastasis-suppressive role for these miRNAs during cancer progression, miR-483 and miR-551a both independently displayed significantly reduced expression levels in human liver metastases relative to primary colon cancers (Figures 6A and 6B; $p = 0.02$ for miR-483 and $p = 0.001$ for miR-551a; $N = 66$). CKB (Figure 6C; $p = 0.05$, $N = 233$) and SLC6a8 (Figure 6D; $p < 0.001$, $N = 233$) expression levels were significantly higher in liver metastases relative to primary tumors in an independent publicly available gene expression data set, suggesting a selection for enhanced expression of CKB and SLC6a8 during progression. We also constructed a tissue microarray from a collection of primary colorectal tumors and liver metastases that were surgically resected from patients at Weill-Cornell

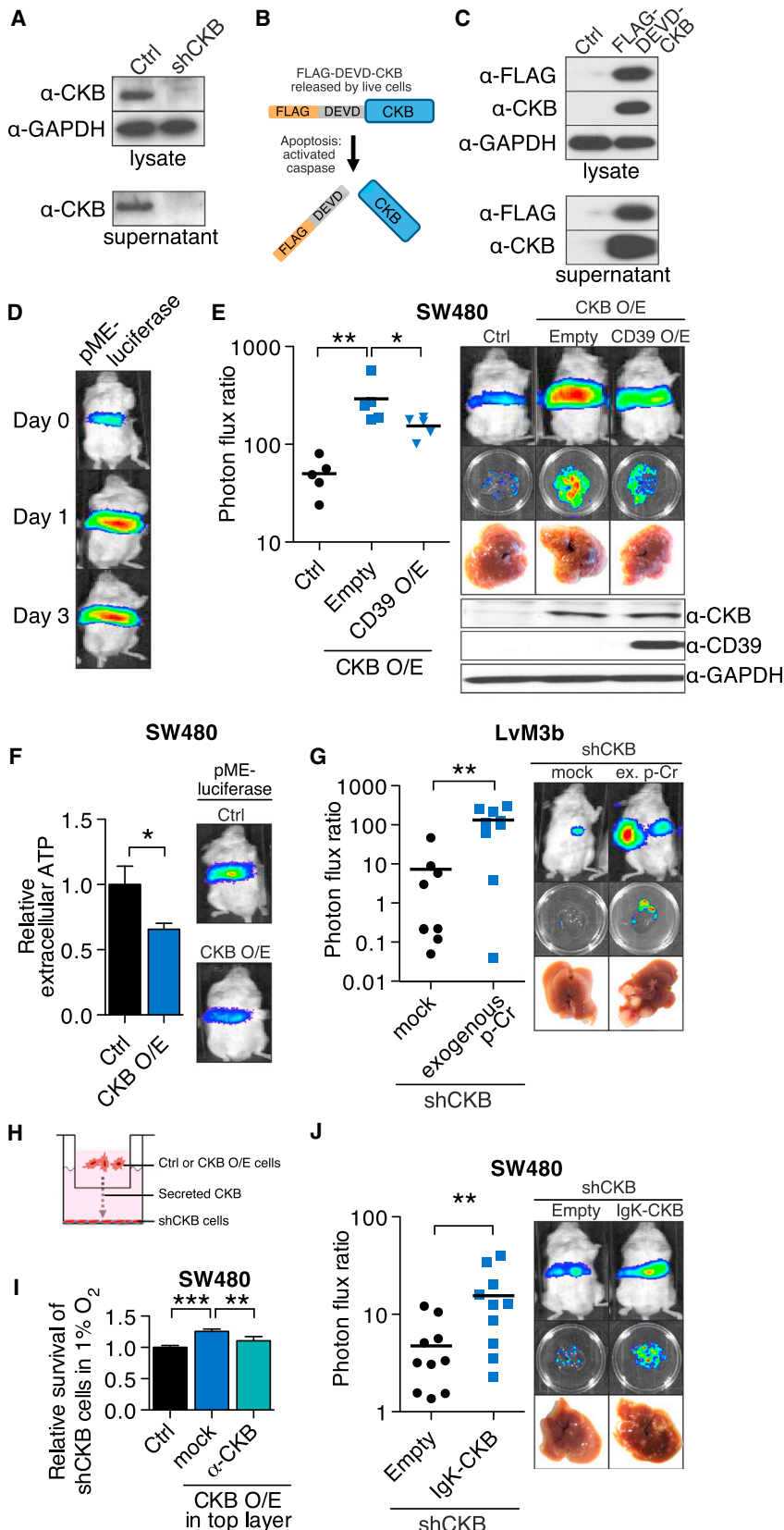


Figure 4. CKB Is Secreted by Colorectal Cancer Cells and Promotes Malignant Conversion of Extracellular ATP and Liver Creatine to Phosphocreatine to Enhance Metastasis

(A) Extracellular and intracellular CKB protein levels in control and Lvm3b cells depleted of CKB through RNAi.

(B) FLAG-tagged CKB with a caspase 3/7 recognition site linker. The FLAG-DEVD-CKB has a FLAG-tag linked to the N terminus of CKB by a linker containing a caspase 3/7 recognition motif (DEVD-amino sequence). Caspase activation in apoptotic cells will result in cleavage of linker and release of FLAG-tag.

(C) Western blot of FLAG-DEVD-CKB overexpressing cells demonstrate release of CKB by non-apoptotic cells into the extracellular space.

(D) Bioluminescent imaging of immunodeficient mice injected with SW480 cells expressing pME-Luc for detection of extracellular ATP (n = 5).

(E) Liver metastasis by 5×10^5 SW480 cells overexpressing CKB with concomitant overexpression of CD39.

(F) Relative extracellular ATP levels in CKB overexpressing cells. Control and CKB overexpressing pME-Luc SW480 cells were injected into mice (n = 5).

(G) Liver metastasis by CKB-depleted Lvm3b cells in mice implanted with an osmotic pump releasing phosphocreatine into the portal circulation.

(H) Scheme for coculture experiment. CKB-knockdown (5×10^4) cells were cultured on the bottom of 24-well plates, while control or CKB-overexpressing cells were plated onto Boyden chambers above CKB-knockdown cells with pores for exchange of metabolites and proteins. Cells at the bottom of the well were counted after 4 days in hypoxia.

(I) Relative survival of CKB-knockdown cells in 1% oxygen when cocultured with control, CKB-overexpressing cells or with CKB-overexpressing cells in the presence of a neutralizing antibody (n = 4).

(J) Liver metastasis by endogenous CKB-knockdown SW480 cells overexpressing a secreted form of CKB. Error bars represent SEM; all p values are based on one-sided Student's t tests, or where appropriate, Mann-Whitney test for non-Gaussian distribution.

*p < 0.05; **p < 0.01; ***p < 0.001. See also Figure S4.

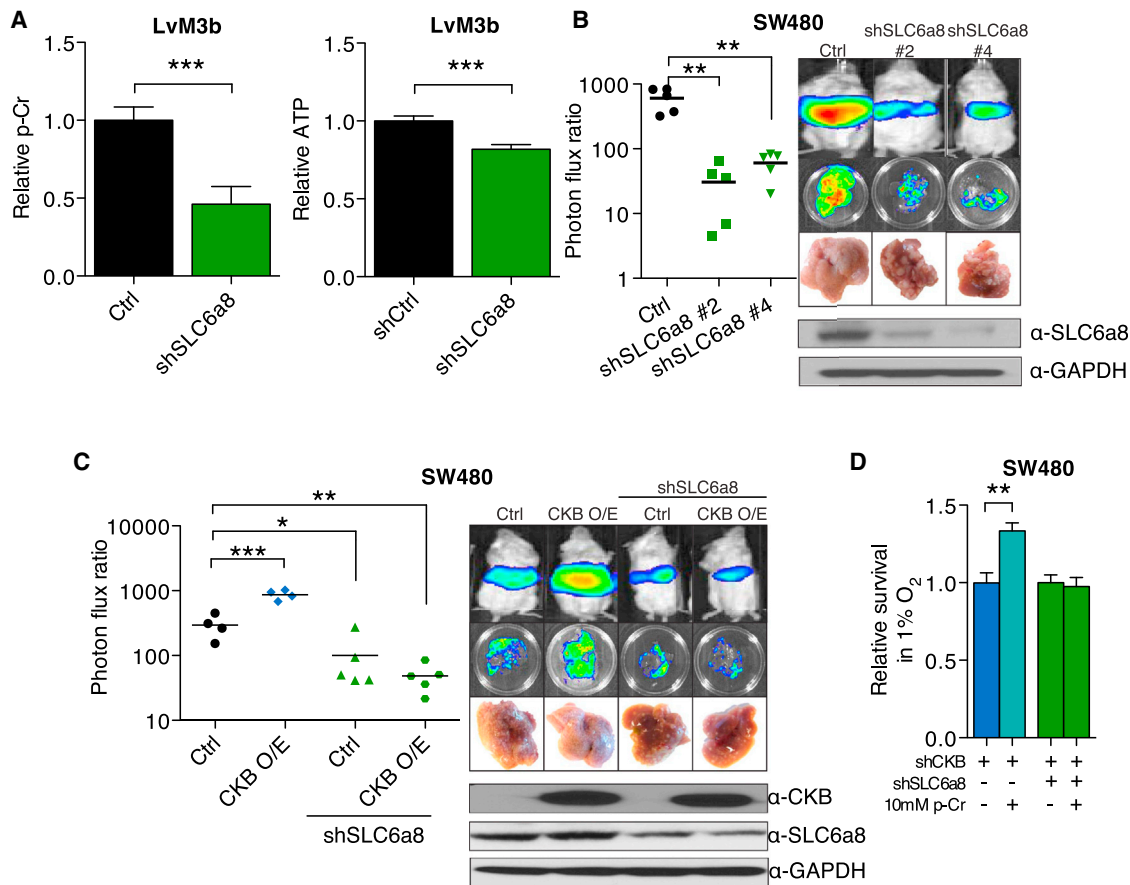


Figure 5. SLC6a8 Modulates CKB-Mediated Colon Cancer Metastasis through Modulation of Intracellular Phosphocreatine Levels

(A) Relative intracellular phosphocreatine and ATP levels in LvM3b cells expressing a control shRNA or shRNA targeting SLC6a8 (n = 4). (B) Liver metastasis by 5×10^5 LvM3b cells expressing two independent short hairpins targeting SLC6a8 (n = 5). (C) Liver metastasis by SW480 cells overexpressing CKB with and without SLC6a8 depletion (n > 4). (D) In vitro survival of LvM3b cells depleted of CKB, SLC6a8 with and without phosphocreatine supplementation in hypoxia (n = 3). Error bars represent SEM; all p values are based on one-sided Student's t tests, or where appropriate, Mann-Whitney test for non-Gaussian distribution. *p < 0.05; **p < 0.001; ***p < 0.0001. See also Figure S5 and Table S4.

Medical Center and immunohistochemically stained for CKB and SLC6a8 expression. CKB (Figure 6E, $p < 0.05$, N = 92) and SLC6a8 (Figure 6F, $p < 0.001$, N = 88) protein expression levels were found to be elevated in liver metastases relative to primary tumors of patients. These findings are consistent with, and suggest the pathophysiological basis for, previous studies revealing elevated expression levels of CKB in advanced stage cancer (Wallimann and Hemmer, 1994) and reveal significant association between the components of this multi-miRNA network and colon cancer progression.

miR-483-5p, miR-551a, and CKB Modulation Provides Clinical Benefit

We sought to investigate the therapeutic potential of targeting this clinically relevant miRNA regulatory network. We first tested the ability of adeno-associated virus to transduce colon cancer cells in vitro and detected viral genomic DNA (gDNA) in colon cancer cells even with low multiplicity of infection (Figure S6A). Injection of mice bearing macroscopic hepatic metastases with

adeno-associated virus revealed that adeno-associated virus was able to infect colon cancer metastases in vivo (Figure S6B). We next injected mice with 5×10^5 highly metastatic LvM3b cells and 24 hr later administered a single intravenous dose of adeno-viral-associated virus (AAV) encoding miR-483-5p and miR-551a from a single transcript. Surprisingly, a single therapeutic dose of AAV delivering both miRNAs dramatically and significantly reduced metastatic colonization (>5-fold; Figure 6G). Therapeutic efficacy was also seen in mice injected with SW480 cells (Figure 6H). Importantly, this treatment did not cause any adverse phenotypic outcomes or pathological abnormalities in mice. Moreover, autopsied mice treated with this AAV therapy did not harbor spontaneous tumors. Even mice injected with BEAS-2B immortalized lung epithelial cells that are prone to oncogenic transformation did not develop tumors (Figure S6C). We next investigated the therapeutic potential of targeting this metabolic network by determining the impact of small-molecule inhibition of CKB on colon cancer metastasis using cyclocreatine. Therapeutic treatment of mice with cyclocreatine after

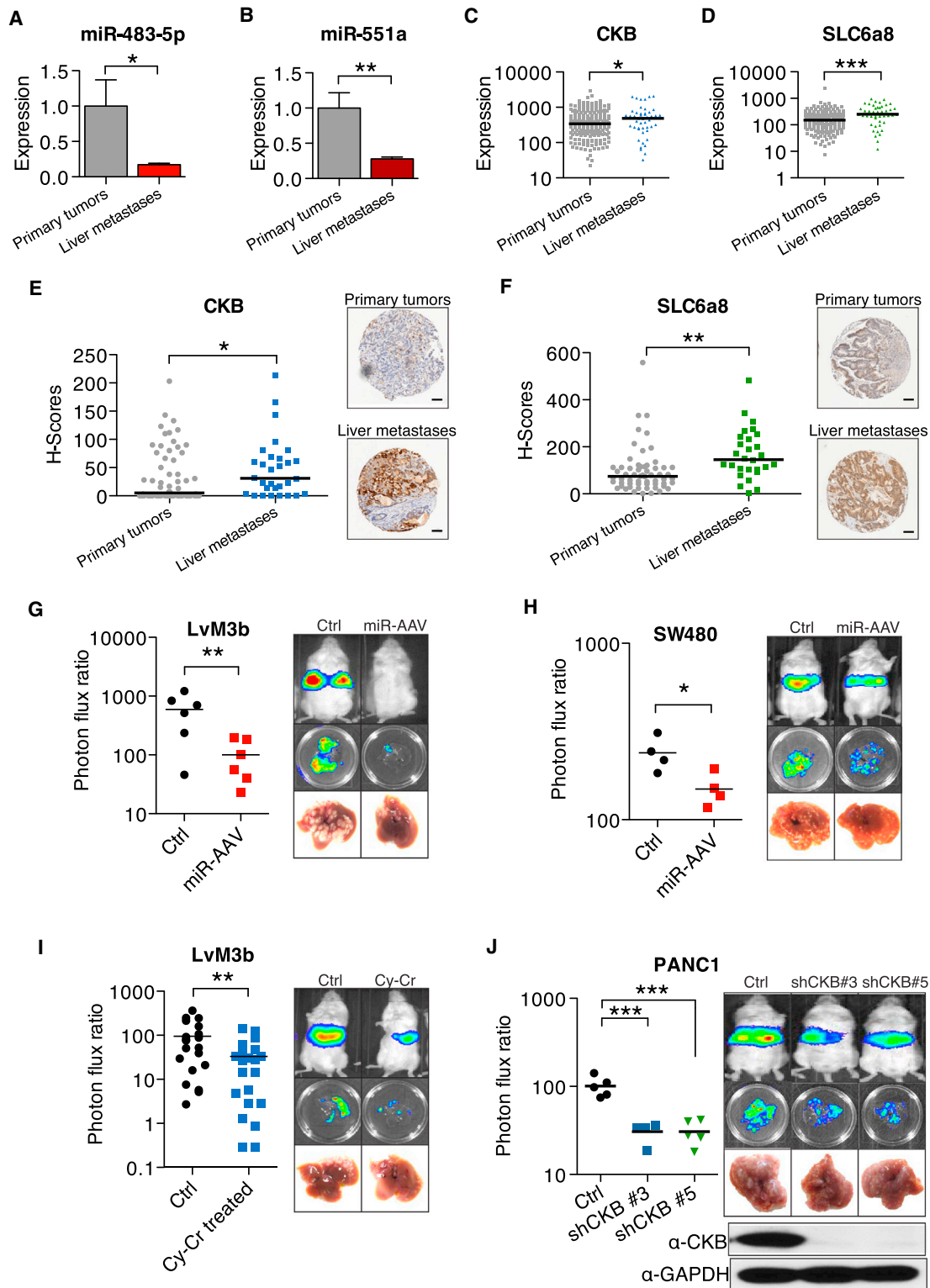


Figure 6. miR-483-5p, miR-551a, CKB, and SLC6a8 Are Clinically Relevant in Independent Cohorts of Patients and Can Be Therapeutically Targeted

(A and B) miR-483-5p and miR-551a levels in 36 primary colorectal adenocarcinomas and 30 liver metastases were quantified by quantitative real-time PCR. (C and D) CKB and SLC6a8 expression from a public microarray data set (GSE41258) comparing primary tumors and liver metastases (N = 233).

(legend continued on next page)

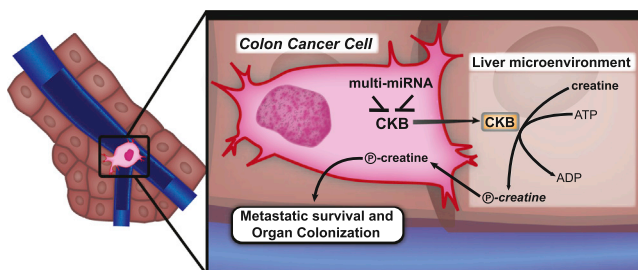


Figure 7. Model for the miR-483-5p, miR-551a, CKB, and SLC6a8 Axis

Disseminated colon cancer cells arrive in the liver microenvironment through the hypoxemic portal circulation. Within the liver microenvironment, they experience hypoxic stress and ATP depletion. Cells that upregulate CKB through loss of miRNAs, release CKB into the extracellular matrix where it converts available creatine and ATP into phosphocreatine that is then taken up by the cell to fuel metastatic survival and subsequent organ colonization. Colon cancer cells with higher levels of CKB also build up a larger pool of intracellular phosphocreatine that acts as a buffer against energetic stress.

colorectal cancer cell inoculation also significantly reduced metastatic colonization, demonstrating proof-of-principle for targeting this kinase as a means of metastasis suppression (Figure 6I). Given that the liver is a common site of metastasis for other gastrointestinal cancers such as pancreatic cancer, we sought to determine if knockdown of components in this pathway could also inhibit metastasis by pancreatic cancer cells. Knockdown of CKB and SLC6a8 in PANC1 cells (a KRAS mutant human pancreatic line) with multiple shRNAs, strongly suppressed their ability to metastasize (Figures 6J and S6G >10-fold). This finding suggests that CKB and SLC6a8, and their associated metabolic pathway, may more broadly govern liver metastasis by other gastrointestinal cancers.

DISCUSSION

Colorectal cancer is diagnosed in over a million patients a year globally, with the majority of advanced stage patients (over 600,000) experiencing liver metastatic progression (Jemal et al., 2011). Using a systematic approach, we have identified two miRNAs that act as suppressors of liver metastatic colonization by colon cancer cells. MiR-483-5p had been recently reported to be oncogenic in lung adenocarcinoma by enhancing invasion and progression and tumor growth in IGF2-dependent sarcoma (Song et al., 2014), while miR-551a has been recently

implicated in suppressing gastric cancer (Li et al., 2012b). Given that miRNAs are widely known to act in a context-specific manner, we had experimentally validated the role of these miRNAs in suppressing liver metastasis by colon cancer cells and we find that these miRNAs suppress a metabolic axis that drives liver colonization by convergent targeting of CKB—a key gene that allows colon cancer cells to expand their phosphocreatine reserves during periods of positive cellular energy balance (Wyss and Kaddurah-Daouk, 2000). Moreover, enhanced phosphocreatine reserves, which can fuel ATP generation, provide great utility during periods of intense energetic requirement such as during metastatic colonization of the liver microenvironment—a hypoxic microenvironment that also contains perivenous hepatocytes that compete with cancer cells for glycolytic substrates (Jungermann and Kietzmann, 1996). We find that colon cancer cells, which enter the liver via the hypoxemic portal circulation, undergo substantial cell death. Cells capable of generating sufficient phosphocreatine reserves are better able to survive the initial phase of hypoxic hepatic colonization. Colon cancer cells that survive the initial selective pressure of hypoxic hepatic dissemination can then activate pathways involved in energy homeostasis and generation (DeBerardinis et al., 2008; Hardie et al., 2012; Inoki et al., 2012; Jeon et al., 2012; Kaelin and McKnight, 2013; Semenza, 2011) and harness additional prometastatic programs for successful competition and further colonization of the liver (Chiang and Massagué, 2008).

As incipient metastatic cells are dependent on CKB-mediated intracellular phosphocreatine generation, which arises from miR-483-5p/miR-551a silencing and CKB overexpression, a selection for subpopulations that exhibit silencing of these miRNAs and consequent induction of CKB would occur. We demonstrate that cells, which overexpress the metabolic kinase CKB, are able to survive and progress within the hepatic parenchyma. However, in the context of energetic stress and ATP limitation, a paradoxical situation would arise wherein intracellular phosphocreatine generation from ATP would comprise a futile cycle. Interestingly, we find that colon cancer cells secrete CKB into the extracellular space, where it catalyzes the ATP-dependent phosphorylation of creatine—yielding phosphocreatine (Figure 7). Extracellular phosphocreatine had been shown to be protective against hypoxic, ischemic, and other energetic insults in neurons and myocardium, with increased phosphocreatine uptake observed in ischemic myocardium (Brustovetsky et al., 2001; Li et al., 2012a; Sharov et al., 1987). Our findings suggest that colon cancer cells can actually generate extracellular phosphocreatine and import it as a means of enhancing energy

(E) CKB expression in primary tumors compared to liver metastases examined through immunohistochemical staining of a tissue microarray (N = 92). Scale bar represents 50 μ m.

(F) SLC6a8 expression in primary tumors compared to liver metastases examined through immunohistochemical staining of above mentioned tissue microarray (N = 88). Scale bar represents 50 μ m.

(G) Liver metastasis in mice injected with 5×10^5 LvM3b cells and treated with a single dose of AAV doubly expressing miR-483-5p and miR-551a 1 day after injection cells (n = 6).

(H) Liver metastasis in mice injected with 5×10^5 SW480 cells and treated with a single dose of AAV doubly expressing miR-483-5p and miR-551a 1 day after injection cells (n = 4).

(I) Liver metastasis in mice injected with 5×10^5 LvM3b cells and treated with cyclocreatine daily for 2 weeks (n > 15).

(J) Liver metastasis by pancreatic cancer cells, PANC1, with knockdown of CKB with two independent shRNA hairpins (n = 5). Error bars represent SEM; all p values are based on one-sided Student's t tests, or where appropriate, Mann-Whitney test for non-Gaussian distribution. *p < 0.05; **p < 0.01; ***p < 0.001.

See also Tables S4 and S5.

stores. Interestingly, we find that colon cancer cells have developed a remarkable adaptive mechanism—secretion of CKB—that enables them to catalytically enhance extracellular phosphocreatine levels from exogenous precursors. The “harvesting” of extracellular metabolites through secretion of CKB by colon cancer cells thus represents a powerful mechanism for survival when malignant cells are highly vulnerable.

While our findings reveal an important prometastatic role for extracellular CKB, we do not rule out the possibility that intracellular CKB could also play a role in cancer progression. One potential intracellular role for CKB might occur during contexts when cancer cells have adequate levels of ATP, such as during primary tumor growth in the colonic epithelium, or subsequent to metastatic tumor expansion and recruitment of functional blood vessels, which could provide adequate oxygenation and glucose for fueling ATP generation. In such contexts, intracellular CKB could allow cancer cells to expand their intracellular buffer of phosphocreatine to be drawn upon during subsequent periods of reduced nutrients. While we have identified one pathway that subserves cancer progression, it is possible that malignant cells conduct other metabolic reactions in the extracellular space that could allow for the extraction of extracellular energy and substrates, or the degradation of toxic metabolic wastes.

Our findings from human primary and metastatic tissue specimens support our model (Figure 7), revealing enhanced expression of CKB and reduced expression of its repressive miRNAs in liver metastases—consistent with a selection for cells with molecular activation of this pathway in the hepatic parenchyma. While our findings have revealed a key pathway that governs colon cancer metastatic colonization of the liver, the predominant site of metastasis by colorectal cancer, and suggest that the *in vivo*-selected metastatic sublines we have derived display the same organotropism for the liver that the majority of human colorectal cancers display, we do not know if this pathway also regulates the colonization of other organs such as the lungs by these cells. We speculate that part of this organotropism may arise from the production of creatine by the liver—the primary organ for creatine biosynthesis. Our findings that CKB and SLC6a8 depletion suppress pancreatic cancer metastasis highlight the possibility that additional gastrointestinal cancer types, which exhibit tropism for the liver, may also utilize this pathway and as such, may be also vulnerable to CKB inhibition therapy.

While the promise of miRNA therapeutics has been great, its actual clinical implementation has been more limited given inadequate stability and delivery of small RNA therapeutics to target tissues. The liver, however, is an exceptional organ in this respect, because miRNAs and RNAi molecules accumulate to higher degrees in the hepatic parenchyma relative to other organs and we have demonstrated adeno-associated virus to be efficient in infecting colon cancer metastases. Additionally, adeno-associated approaches of gene delivery have demonstrated proof-of-concept in human trials (Nathwani et al., 2011). Given these features of the liver, our identification of metastasis suppressor miRNAs in colorectal cancer and our proof-of-principle demonstration of their therapeutic activity have important clinical implications because both nanoparticle-mediated and adeno-

associated viral delivery of these metastasis suppressors could be viable paths clinically. A more conventional path toward targeting this pathway could be the development of highly potent inhibitors of CKB that would act in a similar manner as cyclocreatine, given that highly specific kinase inhibitors can be designed (Dar and Shokat, 2011; Davis and Schlessinger, 2012). Importantly, such inhibitors do not have to be cell-permeable, because they would be targeting extracellular CKB. This could substantially increase the therapeutic index of such compounds. The poor overall efficacy of the current standard-of-care chemotherapy regimen FOLFOX in reducing metastatic relapse rates in high-risk patients necessitates the development and testing of such targeted therapeutic approaches in this prevalent disease.

EXPERIMENTAL PROCEDURES

Animal Studies

All animal work was conducted in accordance with a protocol approved by the Institutional Animal Care and Use Committee (IACUC) at The Rockefeller University. Age-matched male NOD-SCID mice (5- to 6-week-old) were used for organotypic slice culture, intrahepatic colonization, and liver metastasis assays involving LS174T, SW620, WiDR, LvM3a, and LvM3b cell-lines. Age-matched male NOD/SCID gamma male mice (5- to 6-week-old) were used for liver metastasis assays for the SW480 and PANC1 cell lines.

Lenti-miR Library Screening

Cells were transduced with a lentiviral Lenti-miR library (System Biosciences) at a low multiplicity of infection ($\text{moi} < 1$) such that each cell overexpressed a single miRNA. The transduced population was then injected intrahepatically into NOD-SCID mice for *in vivo* selection of miRNAs that when overexpressed, either promoted or suppressed metastatic liver colonization. Genomic DNA PCR amplification and recovery of lenti-viral miRNA inserts was performed on cells prior to injection and from liver nodules according to the manufacturer's protocol. miRNA array profiling allowed for miRNA insert quantification prior to and subsequent to *in vivo* selection.

Organotypic Slice Culture System

Cells to be injected were labeled with cell-tracker red or green (Invitrogen) and inoculated into the livers of NOD-SCID or NOD-SCID gamma mice through intrasplenic injection. The livers were then extracted and cut into 150 μm slices using a McIlwain tissue chopper (Ted Pella) and plated onto organotypic tissue culture inserts (Millipore) and cultured in William's E Medium supplemented with Hepatocyte Maintenance Supplement Pack (Invitrogen). After indicated time periods, the liver slices were fixed in paraformaldehyde and imaged. Extended protocol can be found in the [Extended Experimental Procedures](#).

Adeno-Associated Viral Therapy

miR-483-5p and miR-551a were cloned as a polycistron consisting of both miRNA precursor with flanking genomic sequences in tandem into the BglII and NotI site of scAAV.GFP (Plasmid 21893, Addgene). Adeno-associated virus was packaged, purified and titered by Vector Biolabs. One day after mice were inoculated with colorectal cancer cells, 1×10^{12} AAV viral particles were injected into each mouse through intravenous injection.

Cyclocreatine Treatment of Mice

One day after inoculation of colon cancer cells, mice were injected with 10 mg of cyclocreatine in 300 μl PBS. Treatment was continued daily for 2 weeks until the mice were euthanized.

shRNA and Primer Sequences

shRNA, primers, and cloning sequences are listed in [Tables S4, S5, and S6](#). Additional experimental procedures can be found in the [Extended Experimental Procedures](#).

ACCESSION NUMBERS

The Gene Expression Omnibus (GEO) accession number for the miRNA screen experiments reported in this paper is GSE56320.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and six tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.12.018>.

AUTHOR CONTRIBUTIONS

S.F.T. conceived the project and supervised all research. J.M.L. and S.F.T. wrote the manuscript. J.M.L., A.S., A.N., F.Y.M., and E.W. performed the experiments. P.B.P., Z.Z., and L.S. obtained, curated, and provided access to the clinical samples.

ACKNOWLEDGMENTS

We thank the members of the S.F.T. laboratory and Dr. Siavash Kurdastani for helpful comments and discussion. We thank Dr. Victor Brodsky, Yifang Liu, and Liza Rivera of Weill Cornell Medical College for invaluable help with the tissue microarray and image acquisition. The work was supported by the Irma T. Hirschl/Monique Weill-Caulier Trust, the Starr Cancer Consortium, the Leona M. and Harry B. Helmsley Charitable Trust, and by the NIH Director's New Innovator Award under award number 1DP2OD006506-01. J.M.L. is an A*STAR National Science Scholar. S.F.T. is a cofounder and shareholder of Rgenix and is a member of its scientific advisory board.

Received: April 3, 2014

Revised: September 25, 2014

Accepted: November 19, 2014

Published: January 15, 2015

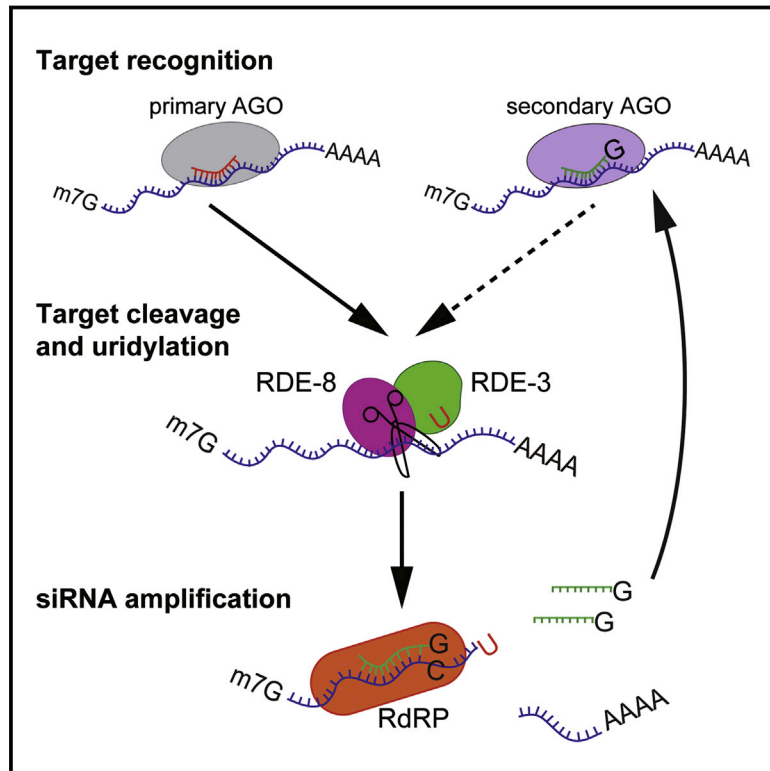
REFERENCES

- Arteel, G.E., Thurman, R.G., Yates, J.M., and Raleigh, J.A. (1995). Evidence that hypoxia markers detect oxygen gradients in liver: pimonidazole and retrograde perfusion of rat liver. *Br. J. Cancer* 72, 889–895.
- Bertout, J.A., Patel, S.A., and Simon, M.C. (2008). The impact of O₂ availability on human cancer. *Nat. Rev. Cancer* 8, 967–975.
- Bruns, C.J., Harbison, M.T., Kuniyasu, H., Eue, I., and Fidler, I.J. (1999). In vivo selection and characterization of metastatic variants from human pancreatic adenocarcinoma by using orthotopic implantation in nude mice. *Neoplasia* 1, 50–62.
- Brustovetsky, N., Brustovetsky, T., and Dubinsky, J.M. (2001). On the mechanisms of neuroprotection by creatine and phosphocreatine. *J. Neurochem.* 76, 425–434.
- Cairns, R.A., Harris, I.S., and Mak, T.W. (2011). Regulation of cancer cell metabolism. *Nat. Rev. Cancer* 11, 85–95.
- Chiang, A.C., and Massagué, J. (2008). Molecular basis of metastasis. *N. Engl. J. Med.* 359, 2814–2823.
- Christofk, H.R., Vander Heiden, M.G., Harris, M.H., Ramanathan, A., Gerszten, R.E., Wei, R., Fleming, M.D., Schreiber, S.L., and Cantley, L.C. (2008). The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature* 452, 230–233.
- Dang, L., White, D.W., Gross, S., Bennett, B.D., Bittinger, M.A., Driggers, E.M., Fantin, V.R., Jang, H.G., Jin, S., Keenan, M.C., et al. (2009). Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* 462, 739–744.
- Dar, A.C., and Shokat, K.M. (2011). The evolution of protein kinase inhibitors from antagonists to agonists of cellular signaling. *Annu. Rev. Biochem.* 80, 769–795.
- Davis, M.J., and Schlessinger, J. (2012). The genesis of Zelboraf: targeting mutant B-Raf in melanoma. *J. Cell Biol.* 199, 15–19.
- DeBerardinis, R.J., Lum, J.J., Hatzivassiliou, G., and Thompson, C.B. (2008). The biology of cancer: metabolic reprogramming fuels cell growth and proliferation. *Cell Metab.* 7, 11–20.
- Fidler, I.J. (1973). Selection of successive tumour lines for metastasis. *Nat. New Biol.* 242, 148–149.
- Gupta, G.P., and Massagué, J. (2006). Cancer metastasis: building a framework. *Cell* 127, 679–695.
- Hardie, D.G., Ross, F.A., and Hawley, S.A. (2012). AMPK: a nutrient and energy sensor that maintains energy homeostasis. *Nat. Rev. Mol. Cell Biol.* 13, 251–262.
- Hickson, J., Ackler, S., Klaubert, D., Bouska, J., Ellis, P., Foster, K., Oleksijew, A., Rodriguez, L., Schlessinger, S., Wang, B., and Frost, D. (2010). Noninvasive molecular imaging of apoptosis in vivo using a modified firefly luciferase substrate, Z-DEVD-aminoluciferin. *Cell Death Differ.* 17, 1003–1010.
- Huddleston, H.G., Wong, K.K., Welch, W.R., Berkowitz, R.S., and Mok, S.C. (2005). Clinical applications of microarray technology: creatine kinase B is an up-regulated gene in epithelial ovarian cancer and shows promise as a serum marker. *Gynecol. Oncol.* 96, 77–83.
- Inoki, K., Kim, J., and Guan, K.L. (2012). AMPK and mTOR in cellular energy homeostasis and drug targets. *Annu. Rev. Pharmacol. Toxicol.* 52, 381–400.
- Jemal, A., Bray, F., Center, M.M., Ferlay, J., Ward, E., and Forman, D. (2011). Global cancer statistics. *CA Cancer J. Clin.* 61, 69–90.
- Jeon, S.M., Chandel, N.S., and Hay, N. (2012). AMPK regulates NADPH homeostasis to promote tumour cell survival during energy stress. *Nature* 485, 661–665.
- Jungermann, K., and Kietzmann, T. (1996). Zonation of parenchymal and non-parenchymal metabolism in liver. *Annu. Rev. Nutr.* 16, 179–203.
- Jungermann, K., and Kietzmann, T. (2000). Oxygen: modulator of metabolic zonation and disease of the liver. *Hepatology* 31, 255–260.
- Kaelin, W.G., Jr., and McKnight, S.L. (2013). Influence of metabolism on epigenetics and disease. *Cell* 153, 56–69.
- Kang, Y., Siegel, P.M., Shu, W., Drobnjak, M., Kakonen, S.M., Cordon-Cardo, C., Guise, T.A., and Massagué, J. (2003). A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* 3, 537–549.
- Kota, J., Chivukula, R.R., O'Donnell, K.A., Wentzel, E.A., Montgomery, C.L., Hwang, H.W., Chang, T.C., Vivekanandan, P., Torbenson, M., Clark, K.R., et al. (2009). Therapeutic microRNA delivery suppresses tumorigenesis in a murine liver cancer model. *Cell* 137, 1005–1017.
- Li, T., Wang, N., and Zhao, M. (2012a). Neuroprotective effect of phosphocreatine on focal cerebral ischemia-reperfusion injury. *J. Biomed. Biotechnol.* 2012, 168756.
- Li, Z., Cao, Y., Jie, Z., Liu, Y., Li, Y., Li, J., Zhu, G., Liu, Z., Tu, Y., Peng, G., et al. (2012b). miR-495 and miR-551a inhibit the migration and invasion of human gastric cancer cells by directly interacting with PRL-3. *Cancer Lett.* 323, 41–47.
- Lillie, J.W., O'Keefe, M., Valinski, H., Hamlin, H.A., Jr., Varban, M.L., and Kaddurah-Daouk, R. (1993). Cyclocreatine (1-carboxymethyl-2-iminoimidazolidine) inhibits growth of a broad spectrum of cancer cells derived from solid tumors. *Cancer Res.* 53, 3172–3178.
- Lujambio, A., and Lowe, S.W. (2012). The microcosmos of cancer. *Nature* 482, 347–355.
- Ma, L., Teruya-Feldstein, J., and Weinberg, R.A. (2007). Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. *Nature* 449, 682–688.
- Markowitz, S.D., and Bertagnoli, M.M. (2009). Molecular origins of cancer: Molecular basis of colorectal cancer. *N. Engl. J. Med.* 361, 2449–2460.
- Meyerhardt, J.A., and Mayer, R.J. (2005). Systemic therapy for colorectal cancer. *N. Engl. J. Med.* 352, 476–487.
- Mingozzi, F., and High, K.A. (2011). Therapeutic in vivo gene transfer for genetic disease using AAV: progress and challenges. *Nat. Rev. Genet.* 12, 341–355.
- Minn, A.J., Gupta, G.P., Siegel, P.M., Bos, P.D., Shu, W., Giri, D.D., Viale, A., Olshen, A.B., Gerald, W.L., and Massagué, J. (2005). Genes that mediate breast cancer metastasis to lung. *Nature* 436, 518–524.

- Nathwani, A.C., Tuddenham, E.G., Rangarajan, S., Rosales, C., McIntosh, J., Linch, D.C., Chowdary, P., Riddell, A., Pie, A.J., Harrington, C., et al. (2011). Adenovirus-associated virus vector-mediated gene transfer in hemophilia B. *N. Engl. J. Med.* **365**, 2357–2365.
- Pellegatti, P., Falzoni, S., Pinton, P., Rizzuto, R., and Di Virgilio, F. (2005). A novel recombinant plasma membrane-targeted luciferase reveals a new pathway for ATP secretion. *Mol. Biol. Cell* **16**, 3659–3665.
- Pellegatti, P., Raffaghello, L., Bianchi, G., Piccardi, F., Pistoia, V., and Di Virgilio, F. (2008). Increased level of extracellular ATP at tumor sites: in vivo imaging with plasma membrane luciferase. *PLoS ONE* **3**, e2599.
- Pencheva, N., and Tavazoie, S.F. (2013). Control of metastatic progression by microRNA regulatory networks. *Nat. Cell Biol.* **15**, 546–554.
- Pencheva, N., Tran, H., Buss, C., Huh, D., Drobnjak, M., Busam, K., and Tavazoie, S.F. (2012). Convergent multi-miRNA targeting of ApoE drives LRP1/LRP8-dependent melanoma metastasis and angiogenesis. *Cell* **151**, 1068–1082.
- Png, K.J., Halberg, N., Yoshida, M., and Tavazoie, S.F. (2012). A microRNA regulon that mediates endothelial recruitment and metastasis by cancer cells. *Nature* **481**, 190–194.
- Rubery, E.D., Doran, J.F., and Thompson, R.J. (1982). Brain-type creatine kinase BB as a potential tumour marker — serum levels measured by radioimmunoassay in 1015 patients with histologically confirmed malignancies. *Eur. J. Cancer Clin. Oncol.* **18**, 951–956.
- Salomons, G.S., van Dooren, S.J., Verhoeven, N.M., Cecil, K.M., Ball, W.S., Degrauw, T.J., and Jakobs, C. (2001). X-linked creatine-transporter gene (SLC6A8) defect: a new creatine-deficiency syndrome. *Am. J. Hum. Genet.* **68**, 1497–1500.
- Semenza, G.L. (2011). Regulation of metabolism by hypoxia-inducible factor 1. *Cold Spring Harb. Symp. Quant. Biol.* **76**, 347–353.
- Semenza, G.L. (2013). HIF-1 mediates metabolic responses to intratumoral hypoxia and oncogenic mutations. *J. Clin. Invest.* **123**, 3664–3671.
- Sharov, V.G., Saks, V.A., Kupriyanov, V.V., Lakomkin, V.L., Kapelko, V.I., Steinschneider, A.Ya, and Javadov, S.A. (1987). Protection of ischemic myocardium by exogenous phosphocreatine. I. Morphologic and phosphorus 31-nuclear magnetic resonance studies. *J. Thorac. Cardiovasc. Surg.* **94**, 749–761.
- Siegel, R., Ma, J., Zou, Z., and Jemal, A. (2014). Cancer statistics, 2014. *CA Cancer J. Clin.* **64**, 9–29.
- Song, Q., Xu, Y., Yang, C., Chen, Z., Jia, C., Chen, J., Zhang, Y., Lai, P., Fan, X., Zhou, X., et al. (2014). miR-483-5p promotes invasion and metastasis of lung adenocarcinoma by targeting RhoGDI1 and ALCAM. *Cancer Res.* **74**, 3031–3042.
- Stagg, J., and Smyth, M.J. (2010). Extracellular adenosine triphosphate and adenosine in cancer. *Oncogene* **29**, 5346–5358.
- Talmadge, J.E., and Fidler, I.J. (2010). AACR centennial series: the biology of cancer metastasis: historical perspective. *Cancer Res.* **70**, 5649–5669.
- Tavazoie, S.F., Alarcón, C., Oskarsson, T., Padua, D., Wang, Q., Bos, P.D., Gerald, W.L., and Massagué, J. (2008). Endogenous human microRNAs that suppress breast cancer metastasis. *Nature* **451**, 147–152.
- Wallimann, T., and Hemmer, W. (1994). Creatine kinase in non-muscle tissues and cells. *Mol. Cell. Biochem.* **133–134**, 193–220.
- Wallimann, T., Wyss, M., Brdiczka, D., Nicolay, K., and Eppenberger, H.M. (1992). Intracellular compartmentation, structure and function of creatine kinase isoenzymes in tissues with high and fluctuating energy demands: the ‘phosphocreatine circuit’ for cellular energy homeostasis. *Biochem. J.* **281**, 21–40.
- Wheaton, W.W., and Chandel, N.S. (2011). Hypoxia. 2. Hypoxia regulates cellular metabolism. *Am. J. Physiol. Cell Physiol.* **300**, C385–C393.
- Wyss, M., and Kaddurah-Daouk, R. (2000). Creatine and creatinine metabolism. *Physiol. Rev.* **80**, 1107–1213.

A Ribonuclease Coordinates siRNA Amplification and mRNA Cleavage during RNAi

Graphical Abstract



Authors

Hsin-Yue Tsai, Chun-Chieh G. Chen, ..., Ming-Daw Tsai, Craig C. Mello

Correspondence

craig.mello@umassmed.edu

In Brief

Although Argonaute proteins are ribonuclease homologs, their enzymatic activity is not required for RNAi in *C. elegans*. Tsai et al. identify RDE-8 as a conserved endoribonuclease that promotes cleavage and uridylation of RNAi targets, which are then used as templates for siRNA amplification. The findings suggest how Argonautes promote both mRNA silencing and the amplification of the silencing signal without directly cleaving their mRNA targets.

Highlights

- RDE-8 encodes a Zc3h12a-like endoribonuclease required for RNAi
- The Argonaute RDE-1 recruits RDE-8 to mRNA targets
- RDE-8 promotes siRNA amplification in vitro and in vivo
- siRNA initiation correlates with target mRNA uridylation

Accession Numbers

GSE59300



A Ribonuclease Coordinates siRNA Amplification and mRNA Cleavage during RNAi

Hsin-Yue Tsai,^{1,3} Chun-Chieh G. Chen,¹ Darryl Conte, Jr.,¹ James J. Moresco,⁴ Daniel A. Chaves,^{1,5} Shohei Mitani,⁶ John R. Yates III,⁴ Ming-Daw Tsai,³ and Craig C. Mello^{1,2,*}

¹RNA Therapeutics Institute, University of Massachusetts Medical School, Worcester, MA 01605, USA

²Howard Hughes Medical Institute, University of Massachusetts Medical School, Worcester, MA 01605, USA

³Institute of Biological Chemistry, Academia Sinica, Nankang, Taipei 115, Taiwan

⁴Department of Chemical Physiology, The Scripps Research Institute, La Jolla, CA 92037, USA

⁵Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Avenida Professor Egas Moniz, 1649-028 Lisboa, Portugal

⁶CREST, Japan Science and Technology Agency and Department of Physiology, Tokyo Women's Medical University School of Medicine, Tokyo 162-8666, Japan

*Correspondence: craig.mello@umassmed.edu

<http://dx.doi.org/10.1016/j.cell.2015.01.010>

SUMMARY

Effective silencing by RNA-interference (RNAi) depends on mechanisms that amplify and propagate the silencing signal. In some organisms, small-interfering RNAs (siRNAs) are amplified from target mRNAs by RNA-dependent RNA polymerase (RdRP). Both RdRP recruitment and mRNA silencing require Argonaute proteins, which are generally thought to degrade RNAi targets by directly cleaving them. However, in *C. elegans*, the enzymatic activity of the primary Argonaute, RDE-1, is not required for silencing activity. We show that RDE-1 can instead recruit an endoribonuclease, RDE-8, to target RNA. RDE-8 can cleave RNA in vitro and is needed for the production of 3' uridylated fragments of target mRNA in vivo. We also find that RDE-8 promotes RdRP activity, thereby ensuring amplification of siRNAs. Together, our findings suggest a model in which RDE-8 cleaves target mRNAs to mediate silencing, while generating 3' uridylated mRNA fragments to serve as templates for the RdRP-directed amplification of the silencing signal.

INTRODUCTION

RNA interference (RNAi) is an ancient gene-silencing mechanism that employs evolutionarily conserved ribonuclease proteins called Argonautes. Argonautes achieve sequence-specific targeting through association with small RNA guides of ~20–30 nucleotides (for review, see Ghildiyal and Zamore, 2009). Pathways related to RNAi are as diverse as the organisms in which they are found and regulate a remarkable array of biological phenomena (Castel and Martienssen, 2013; Conine et al., 2013; Seth et al., 2013).

In *C. elegans*, RNAi triggered by foreign double-stranded RNA (dsRNA) (referred to herein as exo-RNAi) is a two-step Argonaute

response (Yigit et al., 2006). The primary Argonaute RDE-1 is loaded with small interfering RNAs (siRNAs) processed from dsRNA by the ribonuclease-III-related enzyme Dicer (DCR-1). Target recognition by RDE-1/siRNA complexes initiates the amplification of antisense secondary siRNAs, which are synthesized de novo by RNA-dependent RNA polymerase (RdRP) and are primarily 22 nt with a 5'-triphosphorylated guanosine (22G-RNAs; Gu et al., 2009; Pak and Fire, 2007; Sijen et al., 2001, 2007). Secondary siRNAs are loaded onto a family of worm-specific Argonautes (WAGOs), which lack catalytic-site metal-coordinating residues and thus mediate silencing through an unknown mechanism (Yigit et al., 2006). WAGOs include cytoplasmic and nuclear members that also function in multiple endogenous small RNA pathways to silence transposons, cryptic or aberrant genes, and foreign sequences (Gu et al., 2009; Guang et al., 2008, 2010; Shirayama et al., 2012; Yigit et al., 2006).

Endogenous small RNA pathways in *C. elegans* can be classified by their dependence on a primary Argonaute. For example, the Piwi ortholog PRG-1 uses genomically encoded piRNAs (21U-RNAs; Batista et al., 2008; Das et al., 2008; Ruby et al., 2006) to recognize targets with incomplete base-pairing complementarity and initiate a stable and heritable mode of epigenetic silencing known as RNAe (Ashe et al., 2012; Bagijn et al., 2012; Buckley et al., 2012; Lee et al., 2012; Shirayama et al., 2012). The maintenance of RNAe does not require PRG-1 activity but rather depends on RdRPs and both nuclear and cytoplasmic WAGOs, as well as chromatin factors (Ashe et al., 2012; Buckley et al., 2012; Lee et al., 2012; Shirayama et al., 2012). How the small RNA amplification machinery recognizes RNAe targets to maintain 22G-RNA levels at each generation remains unknown.

The ERI (for enhanced RNAi; Kennedy et al., 2004) pathway is a two-step Argonaute pathway that directly competes with the exo-RNAi pathway for available WAGOs (Duchaine et al., 2006; Gent et al., 2010; Vasale et al., 2010; Yigit et al., 2006). The ERI pathway requires both an RdRP (RRF-3) and DCR-1 to generate 26-nt siRNAs with a 5'-monophosphorylated G (Duchaine et al., 2006; Pavelec et al., 2009; Ruby et al., 2006; Vasale et al., 2010). The 26G-RNAs are loaded onto the Argonaute ERGO-1.

Targeting by ERGO-1/26G-RNAs initiates 22G-RNA biogenesis by RdRPs (RRF-1 and EGO-1) and silencing by nuclear and cytoplasmic WAGOs (Gent et al., 2010; Guang et al., 2008; Vasale et al., 2010).

Here, we describe a previously uncharacterized RNAi-deficient mutant, *rde-8*. RDE-8 protein contains a ribonuclease domain known as an N4BP1, YacP Nuclease (NYN) domain (Anantharaman and Aravind, 2006) and is related to the Zc3h12a ribonuclease (Matsushita et al., 2009). We show that RDE-8 is required for the accumulation of two classes of RdRP-dependent small RNAs: RRF-1-dependent 22G-RNAs and RRF-3-dependent 26G-RNAs. We further show that RDE-8 is required for efficient RRF-1 RdRP activity in vitro. Using RNA immunoprecipitation (RIP), we show that RDE-8 associates with target mRNAs during exo-RNAi in an RDE-1-dependent but RdRP-independent manner. We identify RDE-8 homologs and RNAi and transposon-silencing factors as RDE-8-interacting proteins, and we show that RDE-8 localizes to germline Mutator foci. Using 3' rapid amplification of cDNA ends (RACE), we show that RDE-8 promotes the accumulation of target mRNA fragments tailed with untemplated 3' uridine residues. Our findings are consistent with a role for RDE-8 both in mediating mRNA cleavage and promoting amplification of the silencing signal.

RESULTS

rde-8 Encodes a NYN Domain Ribonuclease

In a genetic screen for worms with an RNAi-deficient (Rde) phenotype, we isolated three independent alleles (*ne3309*, *ne3360*, and *ne3361*) of a gene we have named *rde-8*. In addition to the RNAi-deficient phenotype (Figure 1A), we observed a slight developmental delay, increased sensitivity to Orsay virus infection, and germline-transgene desilencing in *rde-8(ne3361)* (Figure S1). Using single-nucleotide polymorphisms and 3-factor analyses, we mapped the *rde-8* gene to a small interval on chromosome IV. Sequencing of candidate genes within this interval revealed that all three alleles harbor the same single-nucleotide (nt) substitution in exon IV of the gene *ZC477.5*, resulting in a nonsense mutation at tryptophan 189 (Figure 1B). Western blot analyses failed to detect the RDE-8 protein in *rde-8(ne3361)* lysates (Figure 1C), suggesting that *ne3361* is a null or strong loss-of-function allele. Two deletion alleles of *rde-8* (*tm2252* and *tm2192*) that remove all or part of exons 4, 5, and 6 (Figure 1B) exhibited the Rde phenotype and failed to complement *rde-8(ne3361)* (data not shown). Finally, an integrated single-copy *gfp::ZC477.5* transgene rescued the Rde phenotype of *rde-8(ne3361)* (Figures 1A and S1). These data identify *ZC477.5* as *rde-8*.

RDE-8 is predicted to encode a 339 amino acid protein homologous to prokaryotic, archaeal, and eukaryotic NYN domain ribonucleases (Figure 1B; Anantharaman and Aravind, 2006). Notably, *gfp::rde-8* transgenes bearing mutations in conserved aspartic acid residues (either D76N alone or D145A and D146A together) that map to the catalytic site of Zc3h12a failed to rescue the Rde phenotype of *rde-8(ne3361)* (Figure 1A and data not shown). Western blot analysis of RDE-8 revealed that the expression of GFP::RDE-8(D76N) protein is comparable to endogenous RDE-8 and wild-type (WT) GFP::RDE-8 (Figure 1C).

These findings suggest that an intact catalytic domain is required for RDE-8 activity.

To directly test whether RDE-8 encodes a ribonuclease, we purified recombinant, histidine-tagged RDE-8(WT) and RDE-8(D76N) proteins by nickel-chelating resin, anion-exchange, and gel-filtration chromatography (Figure 1D). We incubated recombinant RDE-8(WT) or RDE-8(D76N) proteins with an internally labeled 116-nt single-stranded RNA using conditions that support in vitro Zc3h12a nuclease activity (Matsushita et al., 2009). RDE-8(WT) degraded the RNA substrate into variable size fragments, with prominent products of ~20 nt and 30 nt (Figure 1E). These products did not accumulate in reactions with recombinant RDE-8(D76N). Instead, an ~85 nt product accumulated in the RDE-8(D76N) reactions and to much lower levels in RDE-8(WT) reactions. This product could represent an intermediate or, alternatively, the product of a bacterial nuclease contaminating the RDE-8 preparations. These data indicate that RDE-8 encodes an endoribonuclease required for RNAi.

RDE-8 Is Required for the Accumulation of RdRP-Dependent Small RNAs

To explore where RDE-8 functions in the RNAi pathway, we examined small RNA production in mutant and WT *rde-8* transgenic strains exposed to dsRNA targeting the nonessential gene *sel-1* (Figure 2). Northern blot analysis revealed that *sel-1* siRNAs were reduced in *rde-8(ne3361)* relative to WT (Figure 2A) and were rescued in *gfp::rde-8(+)*, but not in *gfp::rde-8(D76N)* transgenic animals (Figures 2A and 2B). The microRNAs *let-7* and *miR-66* were unaffected and serve as loading controls (Figures 2A and 2B). We also cloned and deep sequenced small RNAs from *rde-8(ne3361)* mutants expressing *gfp::rde-8(+)* or *gfp::rde-8(D76N)* and exposed to *sel-1(RNAi)*. Consistent with the northern blot data, we detected secondary siRNAs 5' of the trigger in *gfp::rde-8(+)* worms after 8 hr of exposure to *sel-1* dsRNA, but not in the *gfp::rde-8(D76N)* mutant sample (Figure 2C). By 24 hr, *sel-1* siRNAs throughout the transcript were more abundant in the WT sample than in the *rde-8* mutant sample (Figure 2C). Thus, the RNAi defect of *rde-8* mutants correlates with failure to accumulate RdRP-derived siRNAs.

We also monitored the effect of *rde-8* on the accumulation of endogenous 22G-RNAs (Figures 2D and S2). We found that 22G-RNAs were reduced at least 2-fold in the *rde-8(ne3361)* mutant for 42% ($n = 4,632$) of genes with at least 10 antisense reads per million total reads in WT. The levels of microRNAs and 21U-RNAs were unaffected in *rde-8(ne3361)* mutants (Figures 2A, 2B, and S2). The 22G-RNA defect of *rde-8(ne3361)* was strongly rescued (87% of target genes; $n = 1,938$) by the *gfp::rde-8(+)* transgene but only partially (33.7% of target genes; $n = 1,938$) by the active-site mutant *gfp::rde-8(D76N)* transgene (Figure 2D).

Examining the levels of 22G-RNAs antisense to genes targeted by ERGO-1, WAGO, or CSR-1 (Claycomb et al., 2009; Gu et al., 2009; Vasale et al., 2010), we found that 22G-RNAs antisense to WAGO and ERGO-1 targets were reduced in *rde-8* mutants, whereas 22G-RNAs antisense to CSR-1 targets were mostly unaffected (Figure 2E and Table S1). The ERGO-1- and WAGO-dependent 22G-RNA defects were rescued by

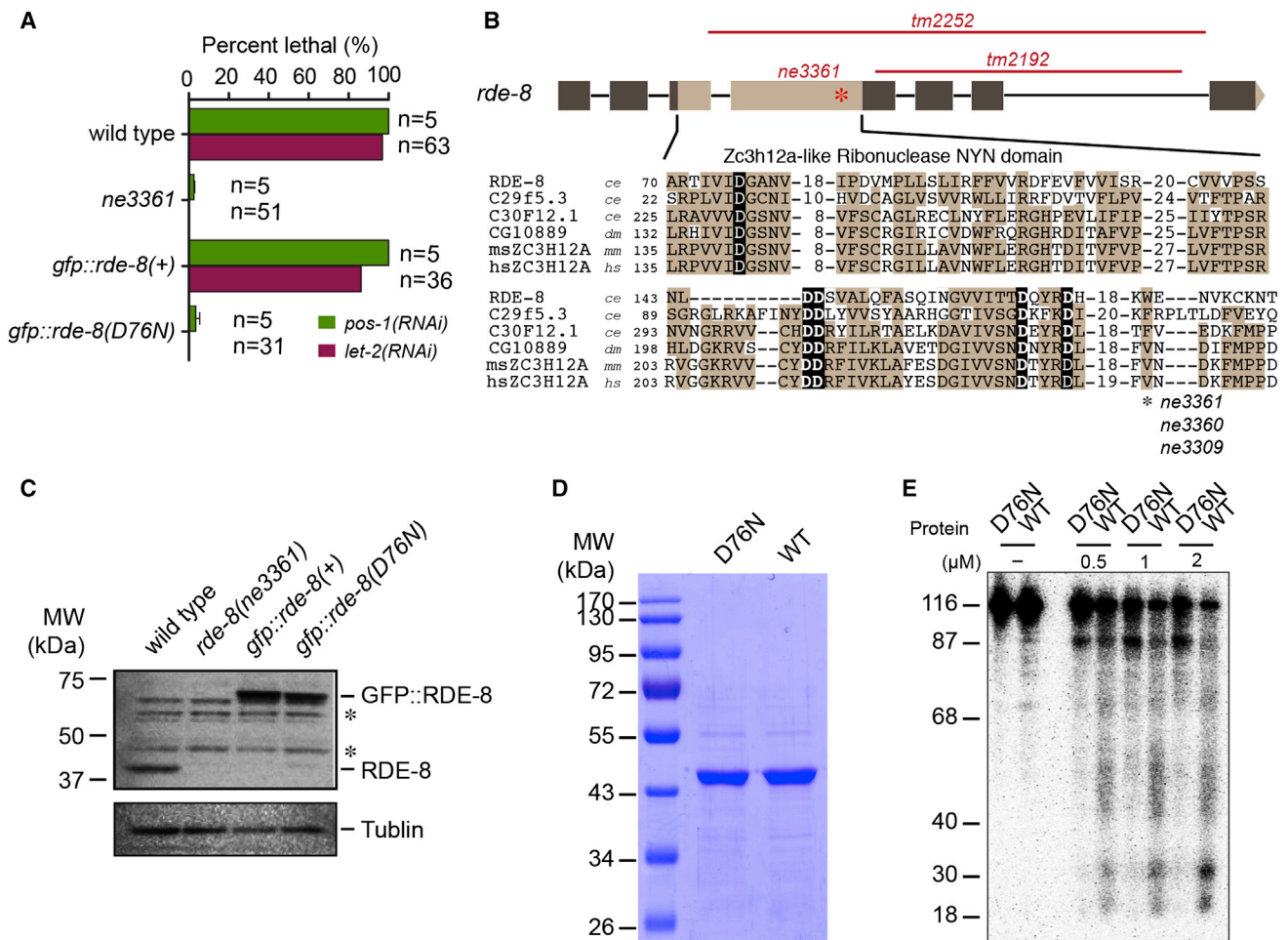


Figure 1. *rde-8* Encodes a Conserved Ribonuclease Required for RNAi

(A) Graphical representation of RNAi sensitivity in WT, *rde-8(ne3361)*, and transgenic strains (as indicated). Percent lethal indicates the mean percentage of *pos-1* dead eggs (green bars) or the percentage of *let-2* ruptured or sterile adults observed (red bars). n, number of animals exposed to RNAi.

(B) Schematic of the *rde-8* locus showing exons (boxes) and intron (lines) with the ribonuclease domain shaded light brown. Deletion (red lines) and nonsense (asterisk) alleles are indicated. The alignment shows *C. elegans* (ce), *Drosophila* (dm), mouse (mm), and human (hs) homologs with conserved residues (shaded brown) and catalytic residues (black background). The asterisk indicates the tryptophan codon (W) mutated in three nonsense alleles.

(C) Immunoblot analysis of RDE-8, GFP::RDE-8, and GFP::RDE-8(D76N) protein expression. Tubulin was probed as a loading control. Asterisks (*) indicate prominent non-specific bands detected by RDE-8 antibody.

(D) Coomassie blue staining of purified WT and D76N recombinant RDE-8 proteins.

(E) Denaturing PAGE analysis of recombinant RDE-8 nuclease activity. RDE-8 protein at different concentrations (indicated) was incubated with a 116-nt *sel-1* RNA (nt 414–529) internally labeled with ³²P-UTP.

See also Figure S1.

the *gfp::rde-8(+)* transgene, but not by the active-site mutant *gfp::rde-8(D76N)* transgene (Figure 2E). WAGO 22G-RNAs dependent on RDE-8 activity included 22G-RNAs that also depend on the PRG-1/piRNA pathway (Table S2; Gu et al., 2009; Lee et al., 2012). RDE-8-dependent 22G-RNAs also included RDE-1/*mir-243*-dependent 22G-RNAs that silence *y47h10a.5* in the soma (Table S1; Corrêa et al., 2010; Gu et al., 2009). Thus the small RNA defects of *rde-8* mutants are consistent with the RNAi and transgene-silencing defects of *rde-8* mutants and suggest that RDE-8 activity is required for the production or accumulation of RdRP-dependent siRNAs that function in the WAGO and ERI silencing pathways.

To ask whether RDE-8 is required for the accumulation of ERGO-1 26G-RNAs, we cloned and deep sequenced 5'-monophosphorylated small RNAs. We found that 26G-RNAs were reduced at least 2-fold in the *rde-8(ne3361)* mutant at 98% (124/126) of ERGO-1 target mRNAs with a minimum of 10 antisense 26G-RNA reads per million total non-structural reads in WT and at least 10-fold at 96% (121/126) of the affected loci (Figure 2F). Interestingly, in the *gfp::rde-8(D76N)* background, 26G-RNAs were reduced by at least 2-fold at only 35% (44/126) of target genes and at least 10-fold at only three of these targets (Figure 2F). Our data suggest that RDE-8 is required for the accumulation of two different classes of RdRP-generated

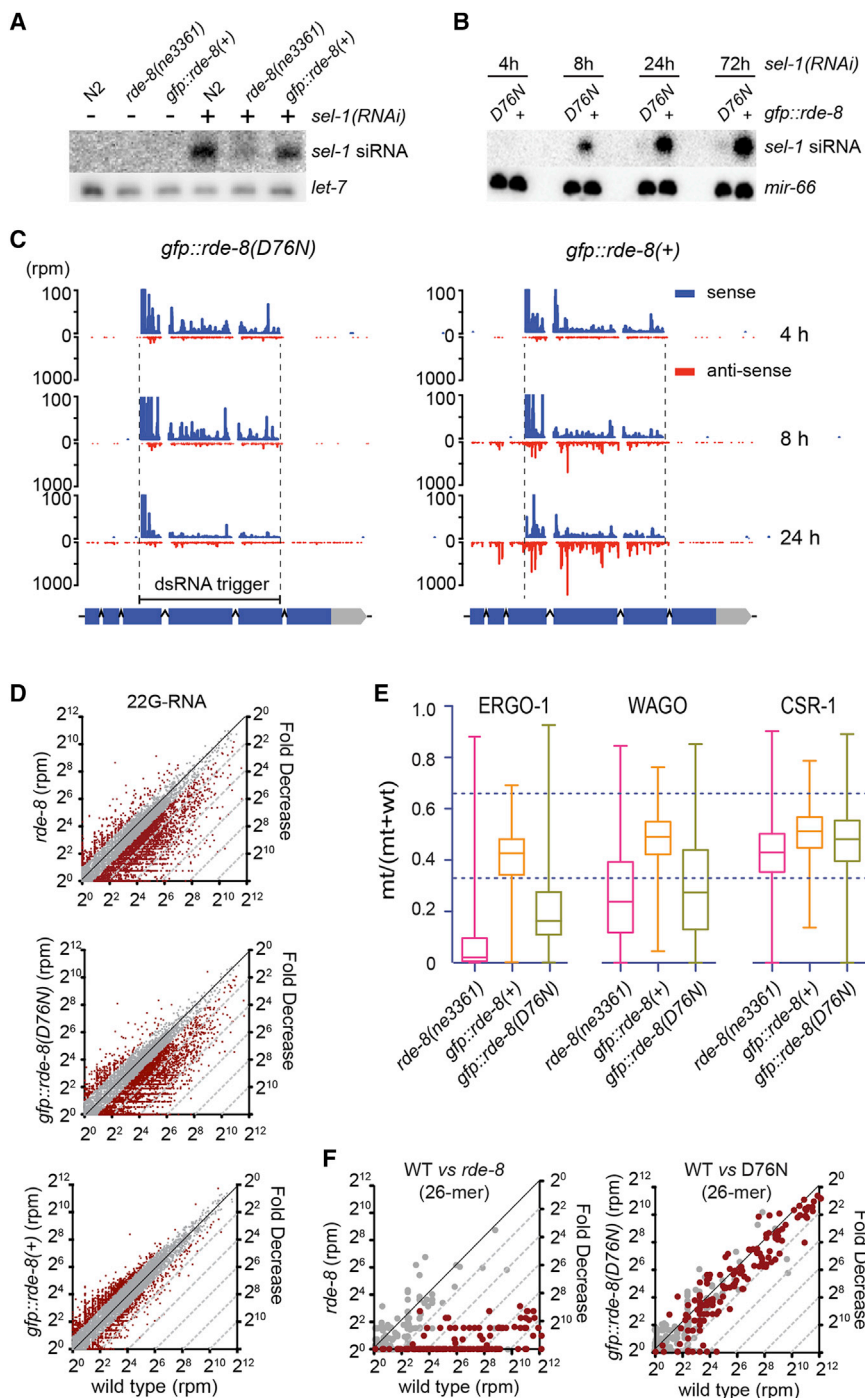


Figure 2. RDE-8 Promotes RdRP-Dependent Small RNA Accumulation

(A and B) Northern blot analyses of antisense *sel-1* siRNAs in WT, *rde-8* mutant, and mutant transgenic strains (as indicated). The probe hybridizes just upstream (5') of the *sel-1* dsRNA trigger region (shown in C). *let-7* and *miR-66* miRNAs were probed as loading controls.

(C) Histograms showing sense (blue) and anti-sense (red) small RNA reads mapping to the *sel-1* gene. Reads were normalized to total non-structural reads. The *sel-1* exons (boxes) and introns (lines) are indicated at bottom; dashed lines delineate the dsRNA trigger region.

(D) Dot plots of endogenous 22G-RNAs targeting annotated genes in *rde-8* (*ne3361*) mutant (top) and mutant transgenic strains (as indicated) *gfp::rde-8(D76N)* (middle) or *gfp::rde-8(+)* (bottom) compared to WT. "rpm" indicates the number of reads per million total reads for a given gene. The black diagonal indicates $x = y$. Dashed lines (gray) demark regions where loci show the indicated fold decrease of 22G-RNA reads compared to WT. Gray dots indicate loci that change less than 2-fold.

(E) Box and whisker plots comparing ERGO-1, WAGO, and CSR-1 pathway 22G-RNAs in *rde-8(ne3361)* mutant (pink) and mutant *gfp::rde-8* transgenic lines (+, orange; D76N, green). The ratio of mutant/(mutant + WT) is shown. The 75th through 25th percentile are boxed, with the median value shown as a horizontal line within the box. Dashed lines indicate 2-fold enrichment (top) and depletion (bottom).

(F) Dot plots of endogenous 26G-RNA levels in *rde-8* (*ne3361*) mutant (left) and mutant *gfp::rde-8* (*D76N*) transgenic animals (right) compared to WT. Red dots represent ERGO-1 targets (Vasale et al., 2010), and gray dots represent loci with non-ERGO-1-associated 26 nt antisense reads. See also Figure S2.

small RNAs, WAGO 22G-RNAs, and ERGO-1 26G-RNAs, but the ribonuclease activity of RDE-8 is not required for 26G-RNA accumulation.

RDE-8 Interacts with ERI/DICER and RNAi/Mutator Pathway Components

To understand how RDE-8 promotes RNAi and 22G-RNA biogenesis, we sought to identify proteins that interact with

WT vs D76N
(26-mer)

Fold Decrease

wild type (rpm)

RDE-8. Using size-exclusion chromatography to examine the molecular weight of RDE-8 complexes in worm lysates, we found that endogenous RDE-8, which has a molecular weight of 38.8 kDa, migrated between 158 kDa (Aldolase) and 440 kDa (Ferritin) in gel filtration analysis (Figure S3). GFP::RDE-8 did not coimmunoprecipitate (co-IP) with endogenous RDE-8 (data not shown), suggesting that the higher-molecular-weight complexes are not composed of RDE-8 multimers. Using multi-dimensional protein identification technology (MudPIT; Chen et al., 2006), we identified several RDE-8-interacting proteins whose loss-of-function phenotypes are similar to those of *rde-8* (Figure 3A and Table S3), including the β -nucleotidyltransferase RDE-3, MUT-15/RDE-5, and the Q/N domain protein MUT-16/RDE-6 (Chen et al., 2005; Vastenhouw et al., 2003; Gu et al., 2009; Zhang et al., 2011).

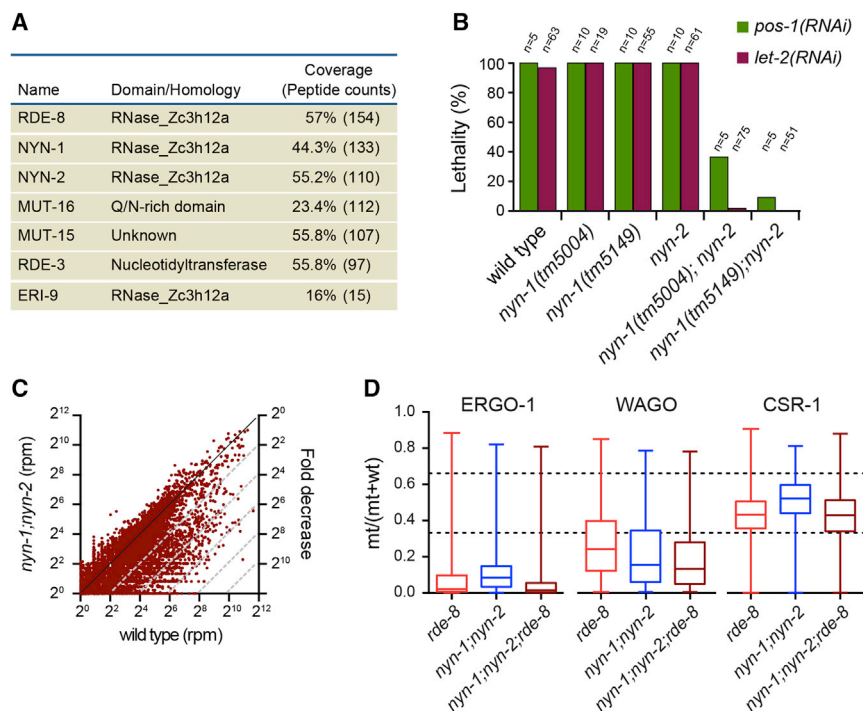


Figure 3. RDE-8 Interacts with Mutator Components and RDE-8 Homologs

(A) Proteins identified by MudPIT analysis of GFP immunoprecipitates from transgenic *gfp::rde-8(+)* worms but not from WT worms that do not express GFP::RDE-8. The percent coverage and total number of peptides are indicated for each RDE-8 interactor.

(B) Graphical representation of RNAi sensitivity of WT or mutant strains (as indicated). Percent lethal indicates the mean percentage of *pos-1* dead eggs (green bars) or the percentage of *let-2* ruptured or sterile adults observed (red bars). n, number of animals exposed to RNAi.

(C) Dot plots (as described in Figure 2D) of endogenous 22G-RNAs targeting annotated genes in *nyn-1(tm5004);nyn-2(tm4844)* double mutants compared to WT.

(D) Box and whisker plots (as described in Figure 2E) comparing ERGO-1, WAGO, and CSR-1 pathway 22G-RNAs in *rde-8(ne3361)* (orange), *nyn-1(tm5004);nyn-2(tm4844)* double mutants (blue), and *rde-8(ne3361);nyn-1(tm5004);nyn-2(tm4844)* triple mutants (brown) relative to WT.

See also Figure S3.

Interestingly, we found that three RDE-8 interactors are homologous to RDE-8, including ERI-9 and two previously unstudied proteins, T23G4.3 and Y87G2A.7, which we have named NYN-1 and NYN-2, respectively. ERI-9 was previously shown to be required for 26G-RNA biogenesis (Pavelec et al., 2009). Consistent with the association of RDE-8 with ERI/Dicer complex components (Duchaine et al., 2006), we found that RDE-8 also co-IPs with the SAP-domain exonuclease ERI-1b (Figure S3; Kennedy et al., 2004), which interacts with both ERI-9 and Dicer and is required for the RdRP-dependent biogenesis of 26G-RNAs (Duchaine et al., 2006; Pavelec et al., 2009; Thivierge et al., 2012).

NYN-1 and NYN-2 are paralogs and more similar to ERI-9 than to other *C. elegans* NYN domain proteins, yet were more highly enriched than ERI-9 in RDE-8 IPs. To test whether NYN-1 and NYN-2 are required for exo-RNAi, we obtained deletion alleles of *nyn-1* (*tm5004* and *tm5149*) and *nyn-2* (*tm4844*). Single mutants were fully sensitive to RNAi in the germline (*pos-1*) and soma (*let-2*), but *nyn-1;nyn-2* double mutants were strongly RNAi deficient in both tissue types (Figure 3B). Thus, NYN-1 and NYN-2 appear to act redundantly in the exo-RNAi pathway.

Consistent with the RNAi defect of *nyn-1;nyn-2* mutants, we found that WAGO-dependent 22G-RNAs and ERI-pathway small RNAs (both 26G-RNAs and 22G-RNAs) were markedly reduced (Figures 3C, 3D, and S3). A triple *nyn-1;nyn-2; rde-8* mutant did not significantly enhance the 22G-RNA defect (Figure S3). Together, our findings suggest that NYN-1 and NYN-2 function with RDE-8 and transposon-silencing factors to promote the biogenesis of RdRP-dependent siRNAs in WAGO- and ERI-dependent silencing pathways.

RDE-8 Localizes to P-Granule-Associated Mutator Foci

The identification of MUT-16/RDE-6, MUT-15/RDE-5, and RDE-3/MUT-2 as RDE-8 interactors suggested that RDE-8 might localize to recently described Mutator foci: perinuclear germline foci that are distinct from, but often adjacent to, germline P granules (Phillips et al., 2012). Indeed, endogenous RDE-8 protein was most abundant in the hermaphrodite or female germline (Figure 4A), and GFP::RDE-8 was primarily observed in the germline cytoplasm and in prominent perinuclear foci associated with nuclear pores in the germline (Figure 4B). Moreover, perinuclear GFP::RDE-8 foci were both fewer in number than and adjacent to P granules identified by RFP::PGL-1 (Figure 4C; Wolke et al., 2007). These data are consistent with the idea that RDE-8 functions along with its interactors MUT-16, MUT-15, and RDE-3 in Mutator foci.

RDE-8 Is Important for Efficient RdRP Activity

RDE-8 is required for RdRP-dependent siRNA accumulation and interacts with several components of Mutator foci, which are thought to be compartments in which RdRP activity promotes siRNA accumulation (Phillips et al., 2012). However, RDE-8 and RdRP interactions were not detected reproducibly in our co-IP studies (data not shown). To ask whether RDE-8 promotes RdRP activity in vitro, we used an assay in which the de novo synthesis of 22G-RNAs is dependent on the RdRP RRF-1, the β -nucleotidyltransferase RDE-3, and a template RNA that is not polyadenylated (Figure 5; Aoki et al., 2007). Consistent with the reduced level of 22G-RNAs observed in *rde-8* mutants, we found that the activity of RdRP was reduced by ~50% in the *rde-8(ne3361)* lysate relative to the WT lysate (Figure 5 and Experimental Procedures). The levels of RRF-1 protein and a

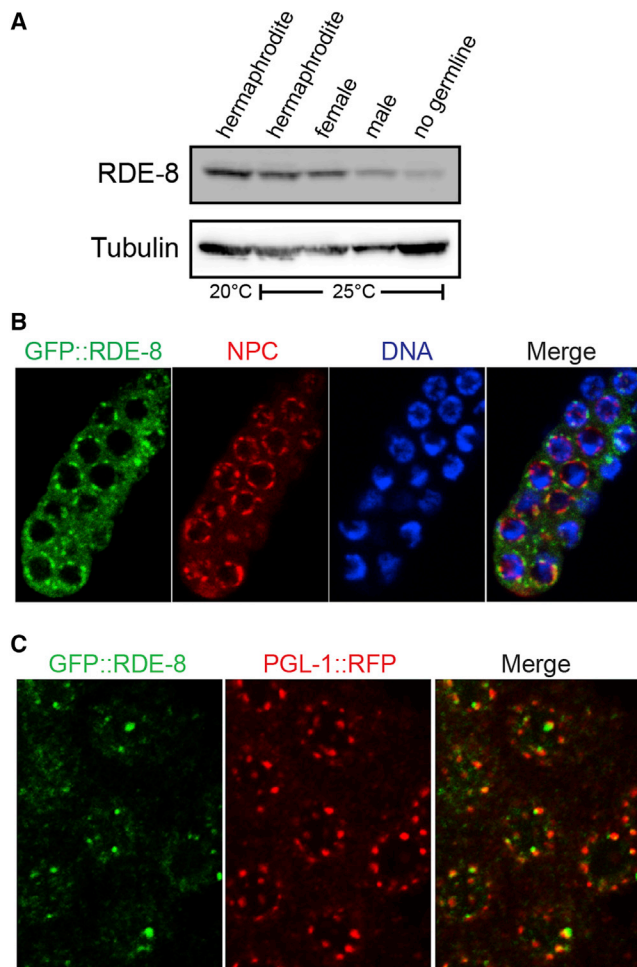


Figure 4. GFP::RDE-8 Localizes to Perinuclear Foci in the Germline
 (A) Immunoblot analysis of RDE-8 protein from WT grown at 20° and 25°C (hermaphrodites), *fem-1(hc17)* grown at 25°C (females), *fog-2(q71)* (males), and from *glp-4(bn2)* animals that lack a germline at 25°C (no germline).
 (B and C) Confocal images of dissected gonads. In (B), gonads expressing GFP::RDE-8 (green) were stained with the MAb414 to detect nuclear pore complex (NPC) proteins (red) and Hoechst to detect DNA (blue). Image overlay at right. In (C), gonads express both GFP::RDE-8 (green) and the constitutive P granule component, RFP::PGL-1 (red). Image overlay at right.

control protein (PRG-1) were similar in the *rde-8(ne3361)* and WT lysates (Figure 5A). RdRP was also less active in a *gfp::rde-8(D76N)* lysate relative to a *gfp::rde-8(+)* lysate (Figure 5B). These findings suggest that RDE-8 is important for efficient RdRP activity.

RDE-8 Interacts with Target mRNA and Requires RDE-1 and Trigger dsRNA

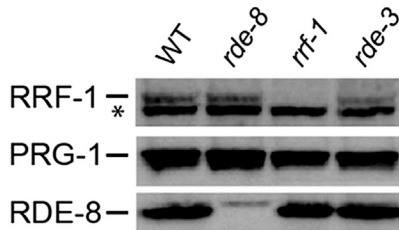
To ask whether RDE-8 interacts with the target mRNA during RNAi, we immunoprecipitated GFP::RDE-8 from worms exposed to dsRNA targeting *sel-1* or a negative control, and then used RT-qPCR to detect regions of the *sel-1* mRNA (Figure 6A). We failed to detect a significant enrichment of *sel-1* mRNA in GFP::RDE-8(WT) IP experiments (Figure 6B). This lack of enrichment

could result from GFP::RDE-8(WT) binding only transiently to the *sel-1* mRNA and then perhaps rapidly cleaving and releasing it. We therefore also tested for RNA binding by the catalytically inactive GFP::RDE-8(D76N). Strikingly, we found that RDE-8(D76N) specifically captured the *sel-1* transcript when animals were exposed to *sel-1* dsRNA (Figure 6A). Interestingly, GFP::RDE-8(D76N) IP enriched similar levels of *sel-1* mRNA from both upstream and downstream of the dsRNA trigger region (regions 1 and 4, Figure 6A), suggesting that GFP::RDE-8(D76N) associates with an intact *sel-1* transcript, one that has not already been cleaved by the primary Argonaute RDE-1 (see below and Discussion).

We next examined the genetic requirement for target mRNA recognition by GFP::RDE-8(D76N). The Argonaute RDE-1 is required for the initiation of RNAi and is loaded with primary siRNAs processed from dsRNA by Dicer (Yigit et al., 2006). Dicer-dependent primary siRNAs are present in *rde-8* mutants (Figure 2C), and, based on affinity capture experiments using 2'-O-methylated RNA oligos, they are loaded onto functional RDE-1 complexes (Figure S4). Consistent with the idea that these primary RDE-1/siRNA complexes are required for RDE-8 to interact with the target, we found that GFP::RDE-8(D76N) failed to capture target mRNA in the *rde-1(ne300)* mutant background (Figure 6B). The ability of RDE-1 to promote RDE-8 binding to the target is likely to be independent of RDE-1 catalytic activity because the catalytic mutant RDE-1(AAA) protein promotes secondary siRNA biogenesis and silencing triggered by dsRNA (Pak et al., 2012; Steiner et al., 2009). As expected, we found that the ability of RDE-1(AAA) to promote *sel-1(RNAi)* is dependent on RDE-8 catalytic activity. *sel-1(RNAi)* dramatically reduced *sel-1* mRNA levels in *rde-1(AAA)* animals, but not in *rde-1(AAA); gfp::rde-8(D76N)* animals (Figure S4). Thus, RDE-8 functions downstream of target recognition by RDE-1.

We next asked whether RdRP is required for RDE-8 to bind target mRNA. To remove RdRP activity, it was necessary to use the double-mutant *rrf-1(pk1417) glp-4(bn2)* background, in which the somatic RdRP *rrf-1* is deleted and the germline RdRP *ego-1* (Smardon et al., 2000) is not expressed due to the absence of germline at 25°C in the temperature-sensitive, germline-deficient mutant *glp-4(bn2)* (Beanan and Strome, 1992). Conditional alleles of *ego-1*, which is an essential gene, do not exist. We found that depletion of RdRP failed to block the association of GFP::RDE-8(D76N) with *sel-1* target mRNA (Figure 6B). Thus, RDE-8 recognizes the target transcript upstream of RdRP and secondary small RNA amplification.

Finally, we asked whether RDE-8 interactors are required for RDE-8 to bind the target. In *mut-15(tm1358)*, *mut-16(tm3748)*, or *nyn-1(tm5004);nyn-2(tm4844)* mutants, enrichment of *sel-1* mRNA by GFP::RDE-8(D76N) RIP was reduced to 22%–42% of the WT level (Figure 6B). By contrast, we found that GFP::RDE-8(D76N) RIP in the *rde-3(ne3370)* mutant background enriched *sel-1* mRNA to 77% of WT levels (Figure 6B). These data suggest that RDE-8 interactors facilitate or stabilize RDE-8 binding to the target mRNA and function together upstream of RdRP. The β -nucleotidyltransferase RDE-3, however, appears to be less important for RDE-8 binding to the target than for RdRP activity, suggesting that RDE-3 may function between RDE-8 and RdRP.



(A) Top: in vitro RdRP activity assayed in WT, *rde-8(ne3361)*, *rrf-1(pk1417)*, or *rde-3(ne3370)* lysates in the presence (+) or absence (–) of in-vitro-transcribed and capped RNA template. “pA” denotes that a polyA stretch was added to the 3’ end of the template. Incubation times are indicated in minutes (min). Oligonucleotide size markers are indicated. “Δ” represents uridylylated template RNA (Aoki et al., 2007). Bottom: Immunoblot analysis of RRF-1, RDE-8, and PRG-1 (loading control) protein levels in lysates used for the RdRP assay.

(B) In vitro RdRP activity assayed in *sel-1* dsRNA-fed, *gfp::rde-8(WT)* and *gfp::rde-8(D76N)* lysates in the presence (+) of in-vitro-transcribed and capped RNA template, as in (A).

Upstream components of the *C. elegans* RNAi machinery must somehow generate mRNA-derived templates for the RdRP-dependent amplification of the silencing signal. In *Tetrahymena*, efficient RdRP recruitment requires 3' uridylation of RNA templates (Lee et al., 2009; Talsky and Collins, 2010). We therefore asked whether target mRNA fragments tailed with untemplated residues accumulate during RNAi in *C. elegans* and, if so, whether or not these products are dependent on RDE-8 activity. To do this, we used 3' RACE to search for mRNA cleavage products in animals exposed to dsRNA targeting the *sel-1* transcript and in control animals not exposed to dsRNA. A 3' linker was ligated to the RNA to provide an anchor for first-strand cDNA synthesis and subsequent PCR reactions. To amplify potential 5' *sel-1* mRNA cleavage products, and not the ingested *sel-1* dsRNA, we amplified each cDNA library using a series of nested *sel-1* mRNA-specific primers to generate 3'-RACE products with 5' ends that lie 40, 30, 20, and 12 nt upstream of the dsRNA target sequence. The nested cDNAs were pooled and amplified, and 3'-RACE products were gel purified and deep sequenced (Figure S5).

To examine the possible relationship between 3' uridylation and RdRP function, we cloned and deep sequenced 22G-RNAs

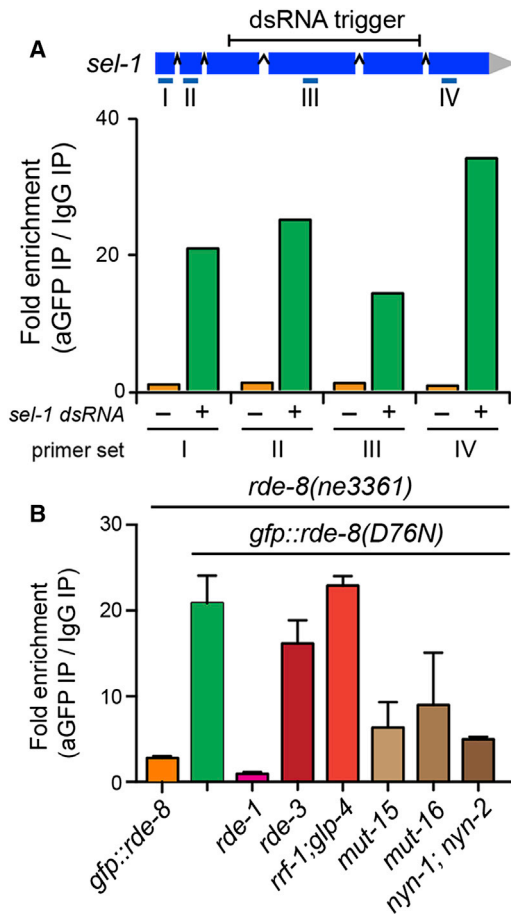


Figure 6. RDE-8 Binds Target mRNA Downstream of RDE-1

(A and B) Bar graphs depicting RT-qPCR results of *sel-1* mRNA levels in GFP immunoprecipitates divided by levels in control IgG precipitates. All strains assayed were *rde-8(ne3361)* and transgenic for *gfp::rde-8(D76N)*, except for one strain that was transgenic for *gfp::rde-8(+)* (orange bar in B). In (A), a schematic diagram of the *sel-1* locus shows the region targeted by dsRNA and the four regions assayed by RT-qPCR. In (A), absence of *sel-1* dsRNA (–) served as a specificity control for each region assayed. In (B), all strains were exposed to *sel-1* dsRNA and were assayed for region II. See also Figure S4.

from each time point in the *gfp::rde-8(+)* and *gfp::rde-8(D76N)* transgenic strains. When we mapped the 3' uridylation and 22G-RNA initiation sites, we observed at least one 22G-RNA peak that initiates 5' and proximal to the major uridylation sites in *gfp::rde-8(WT)*, but not *gfp::rde-8(D76N)*, transgenic worms (Figure 7B). In addition, the 3' uridylation products were observed at 3 hr, whereas the corresponding 22G-RNAs were not detected until 7 hr, suggesting that the accumulation of uridylated *sel-1* fragments precedes the accumulation of the corresponding 22G-RNAs at these sites.

Finally, we examined whether the accumulation of uridylated target mRNA fragments depends on RdRP and RDE-3 activity. Depleting RdRP activity, using an *rrf-1(pk1417)* *gfp-4(bn2)* double-mutant background as described above, we found that *sel-1* mRNA fragments were uridylated proximal to the dsRNA trigger region in *rde-8(ne3361); gfp::rde-8(+)* animals, but uridy-

lation was reduced in *rde-8(ne3361); gfp::rde-8(D76N)* animals and in *rde-3(ne3370)* animals (Figure 7D). In these experiments, the pattern of uridylation differed from that observed in the time course above (perhaps owing to the lack of germline or RdRP). These data suggest that, together, RDE-8 and RDE-3 act upstream of RdRP to promote the uridylation of 5' *sel-1* fragments that could function as RdRP templates.

DISCUSSION

The RdRP-dependent amplification of secondary siRNAs is essential for robust silencing during RNAi in *C. elegans*. Thus, target mRNA destruction must be managed during RNAi so as to preserve mRNA sequences that serve as templates for RdRP amplification. In this study, we have shown that RDE-8 encodes a ribonuclease that interacts with target mRNA downstream of the Argonaute RDE-1 but upstream of RDE-3 and RdRP. RDE-8 forms a complex with RDE-3 in vivo and is required for the uridylation of 5' fragments of the target mRNA and for the amplification of secondary siRNAs by RdRP.

It is well known that a number of viral RdRPs prefer to initiate de novo transcription using GTP (Kao et al., 2001). Furthermore, the *Neurospora* RdRP QDE-1, a homolog of worm RdRPs, prefers to initiate de novo transcription with GTP and to produce 9 to 21 nt small RNAs that are distributed across the template (Makeyev and Bamford, 2002). In *C. elegans*, RdRPs prefer to initiate transcription at a YG motif (as viewed from the antisense strand), where G is the first nucleotide of the 22G-RNA preceded by a pyrimidine (Y), which is similar to the YR motif preferred by RNA polymerase II for transcription initiation, where a purine (R) is the first nucleotide of the transcript (Gu et al., 2012, and references therein).

Our findings suggest a model (Figure 7E) whereby mRNAs are initially recognized but not cleaved by RDE-1. Instead, RDE-1 recruits a complex containing RDE-8, which cleaves the target mRNA exposing a 3' end that can be uridylated by the β -nucleotidyltransferase homolog RDE-3. Thus, RDE-8 cleavage and RDE-3-dependent 3' uridylation of cleavage products may provide a signal or platform to recruit the RdRP complex. RdRP could then, in turn, initiate de novo transcription from internal C nucleotides near the 3' end of the uridylated template. This model is consistent with previous work showing that 22G-RNA amplification is highest proximal to the dsRNA trigger region (Sijen et al., 2001, 2007; Pak and Fire, 2007). RDE-8 may function similarly in the ERI pathway, after ERGO-1 targeting, and may also function along with RDE-3 downstream of WAGO-mediated target recognition.

The nuclease activity of recombinant RDE-8 requires conserved aspartic acid residues that map to the catalytic site of Zc3h12a (Matsushita et al., 2009; Xu et al., 2012). Zc3h12a destabilizes the mRNAs of immune-related factors, including IL-6 and IL-12p40, by directly binding and cleaving 3' UTRs (Matsushita et al., 2009). Zc3h12a was also shown to negatively regulate miRNA expression by cleaving the terminal loop of pre-miRNAs (Suzuki et al., 2011). The CCCH domain of Zc3h12a is required for RNA-binding activity in vitro and for cleavage in vivo, but not in vitro (Suzuki et al., 2011). RDE-8 contains no recognizable RNA-binding domain, and we only detected target binding

when the conserved catalytic residues of RDE-8 were mutated, suggesting a transient or indirect interaction between RDE-8 and the target mRNA. Perhaps consistent with the latter possibility, the interaction with target mRNA was partially dependent on factors that interact with RDE-8 and required the Argonaute RDE-1, which may directly or indirectly recruit the RDE-8 complex to the target mRNA.

Among the factors that interact with RDE-8, we identified three homologs of RDE-8, including ERI-9 and the closely related redundant genes NYN-1 and NYN-2. Previous work has shown that ERI-9 is a component of the ERI-pathway siRNAs expressed in embryos (Duchaine et al., 2006; Pavelec et al., 2009). Our small RNA data indicate that RDE-8, NYN-1, and NYN-2 function along with ERI-9 in the ERI pathway. Remarkably, ERI-9, NYN-1, and NYN-2 lack predicted active-site residues and are thus unlikely to encode functional nucleases. Nevertheless, these factors were required for ERI-pathway 26G-RNA and 22G-RNA biogenesis, and NYN-1 and NYN-2 were also required for RDE-8(D76N) to interact with the target mRNA during RNAi and for RDE-8 to localize to Mutator foci. Together, these results suggest a structural rather than catalytic role for ERI-9, NYN-1, and NYN-2 in the RDE-8 complex. Interestingly, although the catalytic activity of RDE-8 was required for 22G-RNA biogenesis, it was not required for 26G-RNA accumulation. Perhaps RDE-8 and ERI-9 are structurally important for a functional ERI complex, promoting RRF-3-dependent 26G-RNA biogenesis. A structural role for ribonucleases is well documented; the eukaryotic PM/Sci complex, or exosome, for example, is composed of multiple RNasePH family members that lack catalytic capacity (Jain, 2012). Detailed structure-function studies are necessary to sort out the role of these and six other RDE-8 homologs in *C. elegans*. Our findings, along with previous work on Zc3h12a, suggest that members of this conserved nuclease family share ancient and fundamental roles in immunity.

How Does RDE-8 Function during RNAi?

Several studies suggest that direct Argonaute-mediated target mRNA cleavage is not required for mRNA silencing during RNAi (and related pathways) in *C. elegans*. For example, when engineered to contain mutations in conserved metal-coordinating residues needed by other Argonautes for RNA cleavage, RDE-1 and PRG-1 could nevertheless initiate RdRP recruitment and gene silencing in the dsRNA- and piRNA-initiated pathways, respectively (Bagijn et al., 2012; Lee et al., 2012; Pak and Fire, 2007; Shirayama et al., 2012; Steiner et al., 2009). Moreover, all 12 of the downstream WAGO Argonautes, required for silencing in both of these pathways, encode proteins that lack key residues required for target cleavage (Yigit et al., 2006). Similarly, *Rhodobacter sphaeroides* Argonaute is not a functional endonuclease, but promotes the silencing of foreign genetic material (Olovnikov et al., 2013). These observations suggest that, in the absence of endonuclease activity, Argonautes can promote silencing by guiding other nucleases or turnover pathways to their targets (Huntzinger and Izaurralde, 2011; Lim et al., 2014). Our finding that target mRNA cleavage products with untemplated uridine residues accumulate during RNAi in an RDE-8-dependent manner raises the possibility that RDE-8 (or other components of the RDE-8 complex) may provide cleavage activ-

ity important for RNAi. Our findings place RDE-8 downstream of primary Argonautes (RDE-1, ERGO-1, and PRG-1) and upstream or at the same level as RdRP in each of these pathways.

Several recent studies have identified factors that appear to function between RDE-1 and RdRP, including RDE-10, RDE-11, and RDE-12 (Shirayama et al., 2014; Yang et al., 2012, 2014; Zhang et al., 2011). Based on IP experiments presented here and in the aforementioned studies, RDE-8 does not interact with these factors. Furthermore, RDE-10, RDE-11, and RDE-12 are specifically required for WAGO-associated 22G-RNAs dependent on RDE-1 and ERGO-1 (Shirayama et al., 2014; Yang et al., 2012, 2014; Zhang et al., 2011), whereas RDE-8 is more broadly required for WAGO-associated 22G-RNAs. Thus, RDE-8 may function at a distinct step (or multiple steps) in the RNAi pathway.

Finally, RDE-8 also promotes the accumulation of WAGO-associated 22G-RNAs that are independent of known primary Argonautes and thus appear to function in self-enforcing trans-generational silencing pathways. It is tempting to speculate that WAGOs have evolved catalytic mutations so that they do not cleave the target mRNA within the region required to template de novo synthesis of the siRNAs that successfully guided them to the target. Instead, WAGOs may recruit secondary nucleases that cleave the target mRNA 3' of where their guide RNAs engage the target. The β -nucleotidyltransferase homolog RDE-3 could then modify this cleavage product to stabilize it or to recruit RdRP to regenerate the 22G-RNA, thereby propagating self-enforcing silencing signals (Figure 7E). In contrast, perhaps RDE-1 recruits RDE-8 more broadly along the target mRNA, producing multiple cleavage products that can serve to generate new RdRP-derived siRNAs (Figure 7E).

Conclusion

RDE-8 homologs regulate post-transcriptional silencing in innate immune pathways in both worms and mammals. Our findings suggest that RDE-8 homologs might function as nucleases or as structural subunits of silencing complexes that promote 3' uridylation of substrates. Uridylation plays diverse and important roles in small RNA pathways (Lee et al., 2014; Talsky and Collins, 2010). In plants and animals, 3' uridylation promotes the turnover of miRNAs and siRNAs (Ameres et al., 2010; Ibrahim et al., 2010), and miRNA- or siRNA-directed cleavage of mRNA results in 3' uridylation of the 5' cleavage product and rapid mRNA decay (Shen and Goodman, 2004). Moreover, Kim and colleagues have recently shown that 3' uridylation enhances mRNA decay of deadenylated miRNA targets (Lim et al., 2014). Perhaps 3' uridylation of transcripts processed by RDE-8-related nucleases is a conserved signal for post-transcriptional silencing. It will therefore be interesting to learn whether RDE-8-related proteins function broadly in uridylation-dependent pathways that regulate gene expression.

EXPERIMENTAL PROCEDURES

Genetics

C. elegans culture and genetics were performed essentially as described (Brenner, 1974). Unless otherwise noted, the WT strain in this study is the Bristol N2 strain. Alleles used in this study listed by

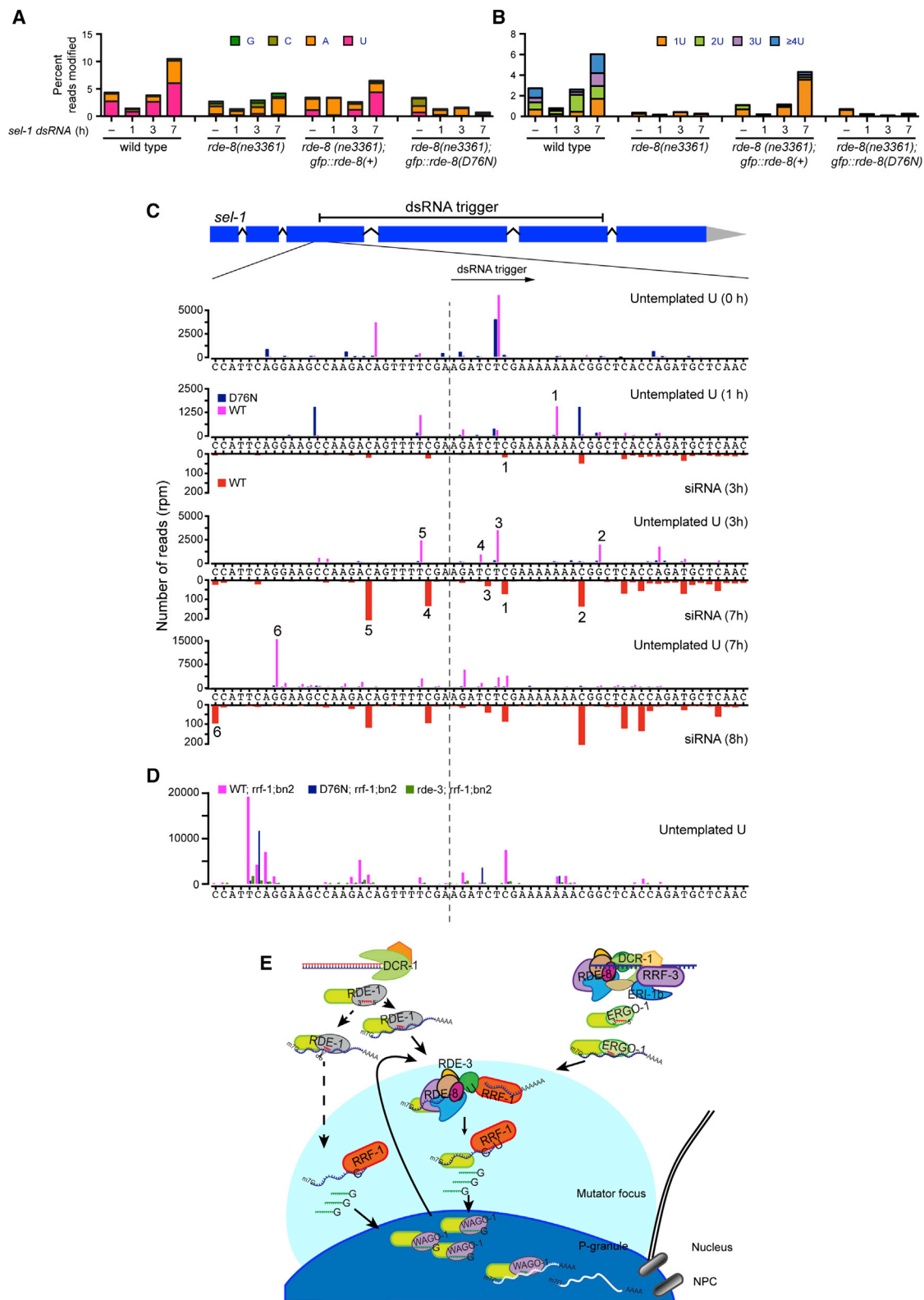


Figure 7. RDE-8 Promotes Target mRNA Cleavage and 3' Uridylation Adjacent to Sites of Secondary siRNA Initiation

(A) Bar graphs showing the percentage of untemplated residues detected by 3' RACE at 3' ends of *sel-1* mRNA fragments during *sel-1*(RNAi). WT or *rde-8(ne3361)* animals were exposed to control dsRNA (–) or *sel-1* dsRNA for 1, 3, or 7 hr.

(legend continued on next page)

chromosome: LGI: *mut-16(tm3748, ne322)*, *rde-3(ne3370)*, *rff-1(pk1417)*, *glp-4(bn2)*, *nyn-2(tm4844)*; LGII: *neSi24[gfp::rde-8(+), cb-unc-119(+)]*, *neSi25[gfp::rde-8(D76N), cb-unc-119(+)]*; LGIV: *rde-8(ne3361, tm2192, tm2252)*, *fem-1(hc17)*, *nyn-1(tm5004, tm5149)*; LGV: *rde-1(ne300)*, *mut-15(tm1358)*, *fog-2(q71)*. The genetic screen and transgenic procedures are detailed in [Extended Experimental Procedures](#).

Recombinant RDE-8 Protein Purification and Ribonuclease Assay

rde-8 cDNAs (WT and D76N) were cloned into a pET expression vector. Expression was induced in BL21(DE3) cells at 16°C overnight. 6His-RDE-8 fusion proteins were extracted in 50 mM Tris (pH 7.5), 50 mM NaCl by sonication. Soluble 6His-RDE-8 fusion proteins were purified by anion exchange (Q-Sepharose), nickel-chelate, and gel-filtration chromatography. Proteins purity was verified by SDS-PAGE. Nuclease assays were performed as described in [Matsushita et al. \(2009\)](#), but the buffer was adjusted to pH 6.0.

RNA Immunoprecipitation

Worms were harvested as adults and extracted in 1 pellet volume of homogenization buffer (25 mM HEPES-KOH (pH 7.5), 10 mM potassium acetate, 2 mM magnesium chloride, 0.1% NP-40, 110 mM potassium chloride, 200 U/ml SUPERaseIn [Ambion]). 20 mg of lysate was incubated with 20 µg of anti-GFP antibody (Wako) or IgG control for 1 hr at 4°C. Immune complexes were captured with Protein A/G-Sepharose beads (Santa Cruz Biotechnology) and washed with homogenization buffer. RNA was extracted with Trizol (MRC, Inc.), and cDNA was prepared using SuperScript III (Life Technologies) and a mixture of *sel-1* and *act-3* primers. The level of *sel-1* mRNA was measured by quantitative PCR relative to *act-3*. Primers are listed in [Extended Experimental Procedures](#).

MudPIT

MudPIT analysis was performed essentially as described ([Conine et al., 2013](#)). See [Extended Experimental Procedures](#) for details.

Small RNA Cloning and Data Analysis

Small RNAs between 18 and 40 nt were gel purified and cloned after treatment with CIP (NEB) and PNK (NEB) or without pretreatment (direct cloning) as described ([Gu et al., 2009](#); [Vasale et al., 2010](#)). Libraries were sequenced either on an Illumina GAI or HiSeq instrument in the UMass Medical School Deep Sequencing Core. Sequences were aligned to the worm genome (WS235) using Bowtie 0.12.9 ([Langmead et al., 2009](#)). Custom Python scripts used to process and analyze the data are available upon request. See [Extended Experimental Procedures](#) for details.

3' RACE Sequencing

Total RNA was extracted from worms fed with bacteria expressing *sel-1* or control dsRNA. RNAs were ligated to the activated 3' linker, miR Linker 1 (IDT). Ligation products were reverse transcribed using a primer specific to the 3' linker sequence. Libraries were generated by nested PCR and sequenced on an Illumina HiSeq 2000 instrument. See [Extended Experimental Procedures](#) for details.

RdRP Assay

The in vitro RdRP assay ([Aoki et al., 2007](#)) was performed four times. Lysates were prepared from synchronized adult worms. Labeled products separated on a 15% polyacrylamide/8 M urea gel were detected and quantified using a

Storm phosphorimager and ImageQuant software (GE Healthcare). The rate of ³²P-UTP incorporation into siRNAs by RdRP was calculated as the slope of time course data (e.g., [Figure 5](#)) fitted to a linear function.

Fluorescence Imaging

Gonads of adult *rde-8(ne3361);gfp::rde-8(+)* animals were extruded by decapitating worms on a slide. Extruded gonads were fixed with formaldehyde, permeabilized by freeze-cracking, and stained as described ([Phillips et al., 2009](#)). Confocal images were acquired using a Zeiss LSM510 Meta Confocal microscope and Zen 2008 (Zeiss) software.

ACCESSION NUMBERS

Illumina data are available from GEO under the accession number GSE59300.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and three tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.01.010>.

ACKNOWLEDGMENTS

We thank members of the Mello lab for great discussion; E. Kittler and the UMass Deep Sequencing Core for Illumina sequencing; P. Furcinitti and the UMass Light Microscopy Core and the Academia Sinica, Institute of Biological Chemistry Imaging & Cell Biology Core facilities for help with confocal microscopy. Some strains were provided by the *Caenorhabditis* Genetics Center, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). H.-Y.T. was supported by a fellowship from the Leukemia and Lymphoma Society (5247-09). J.J.M. and J.R.Y. were supported by the National Center for Research Resources (5P41RR011823), National Institute on Aging (R01AG027463), and National Institute of General Medical Sciences (8P41GM103533). D.A.C. is supported by Fundação para Ciência e a Tecnologia, Portugal (SFRH/BD/17629/2004/H6BM). M.-D.T. is supported by Academia Sinica. C.C.M. is a Howard Hughes Medical Institute Investigator and is supported by NIH grants GM058800 and 1S10RR027052.

Received: July 29, 2014

Revised: October 21, 2014

Accepted: December 23, 2014

Published: January 29, 2015

REFERENCES

- Ameres, S.L., Horwich, M.D., Hung, J.H., Xu, J., Ghildiyal, M., Weng, Z., and Zamore, P.D. (2010). Target RNA-directed trimming and tailing of small silencing RNAs. *Science* 328, 1534–1539.
- Anantharaman, V., and Aravind, L. (2006). The NYN domains: novel predicted RNases with a PIN domain-like fold. *RNA Biol.* 3, 18–27.
- Aoki, K., Moriguchi, H., Yoshioka, T., Okawa, K., and Tabara, H. (2007). In vitro analyses of the production and activity of secondary small interfering RNAs in *C. elegans*. *EMBO J.* 26, 5007–5019.
- Ashe, A., Sapetschnig, A., Weick, E.M., Mitchell, J., Bagijn, M.P., Cording, A.C., Doebley, A.L., Goldstein, L.D., Leirbach, N.J., Le Pen, J., et al. (2012).

(B) Bar graphs, as in (A), showing the percentage of *sel-1* mRNA fragments with 1, 2, 3, or 4+ untemplated uridine residues during *sel-1(RNAi)*.

(C) Sequence alignment of uridylation sites in WT (magenta) and D76N (blue) within the region of *sel-1* just 5' of the dsRNA trigger. 3' RACE products were sequenced at four time points: prior to dsRNA exposure (0 hr) and after 1, 3, or 7 hr (as indicated). The 5' ends of 22G-siRNAs (plotted below the sequence, orange bars) were identified in the *rde-8(ne3361);gfp::rde-8(+)* transgenic strain. Three time points are shown. The *sel-1* exon (box), intron (line) structure, the dsRNA trigger region, and the region of sequence analyzed are indicated in the schematic.

(D) Sequence alignment, as in (C), showing uridylation sites in RdRP and *rde-3* mutants. The *rff-1 glp-4(bn2)* background was used to remove RdRP activity in each strain analyzed, including *rde-8(ne3361);gfp::rde-8(+)* (WT, purple), *rde-8(ne3361);gfp::rde-8(D76N)* (D76N, blue), and *rde-3(ne3370)* (green).

(E) RDE-8 coordinates siRNA amplification and mRNA cleavage during RNAi. See [Discussion](#) for details of the model.

See also [Figure S6](#).

piRNAs can trigger a multigenerational epigenetic memory in the germline of *C. elegans*. *Cell* 150, 88–99.

Bagijn, M.P., Goldstein, L.D., Sapetschnig, A., Weick, E.M., Bouasker, S., Lehrbach, N.J., Simard, M.J., and Miska, E.A. (2012). Function, targets, and evolution of *Caenorhabditis elegans* piRNAs. *Science* 337, 574–578.

Batista, P.J., Ruby, J.G., Claycomb, J.M., Chiang, R., Fahlgren, N., Kasschau, K.D., Chaves, D.A., Gu, W., Vasale, J.J., Duan, S., et al. (2008). PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in *C. elegans*. *Mol. Cell* 31, 67–78.

Beanan, M.J., and Strome, S. (1992). Characterization of a germ-line proliferation mutation in *C. elegans*. *Development* 116, 755–766.

Brenner, S. (1974). The genetics of *Caenorhabditis elegans*. *Genetics* 77, 71–94.

Buckley, B.A., Burkhart, K.B., Gu, S.G., Spracklin, G., Kershner, A., Fritz, H., Kimble, J., Fire, A., and Kennedy, S. (2012). A nuclear Argonaute promotes multigenerational epigenetic inheritance and germline immortality. *Nature* 489, 447–451.

Castel, S.E., and Martienssen, R.A. (2013). RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nat. Rev. Genet.* 14, 100–112.

Chen, C.C., Simard, M.J., Tabara, H., Brownell, D.R., McCollough, J.A., and Mello, C.C. (2005). A member of the polymerase beta nucleotidyltransferase superfamily is required for RNA interference in *C. elegans*. *Curr. Biol.* 15, 378–383.

Chen, E.I., Hewel, J., Felding-Habermann, B., and Yates, J.R., 3rd. (2006). Large scale protein profiling by combination of protein fractionation and multi-dimensional protein identification technology (MudPIT). *Mol. Cell. Proteomics* 5, 53–56.

Claycomb, J.M., Batista, P.J., Pang, K.M., Gu, W., Vasale, J.J., van Wolfswinkel, J.C., Chaves, D.A., Shirayama, M., Mitani, S., Ketting, R.F., et al. (2009). The Argonaute CSR-1 and its 22G-RNA cofactors are required for holocentric chromosome segregation. *Cell* 139, 123–134.

Conine, C.C., Moresco, J.J., Gu, W., Shirayama, M., Conte, D., Jr., Yates, J.R., 3rd, and Mello, C.C. (2013). Argonautes promote male fertility and provide a paternal memory of germline gene expression in *C. elegans*. *Cell* 155, 1532–1544.

Corrêa, R.L., Steiner, F.A., Berezikov, E., and Ketting, R.F. (2010). MicroRNA-directed siRNA biogenesis in *Caenorhabditis elegans*. *PLoS Genet.* 6, e1000903.

Das, P.P., Bagijn, M.P., Goldstein, L.D., Woolford, J.R., Lehrbach, N.J., Sapetschnig, A., Buhecha, H.R., Gilchrist, M.J., Howe, K.L., Stark, R., et al. (2008). Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the *Caenorhabditis elegans* germline. *Mol. Cell* 31, 79–90.

Duchaine, T.F., Wohlschlegel, J.A., Kennedy, S., Bei, Y., Conte, D., Jr., Pang, K., Brownell, D.R., Harding, S., Mitani, S., Ruvkun, G., et al. (2006). Functional proteomics reveals the biochemical niche of *C. elegans* DCR-1 in multiple small-RNA-mediated pathways. *Cell* 124, 343–354.

Gent, J.I., Lamm, A.T., Pavelec, D.M., Maniar, J.M., Parameswaran, P., Tao, L., Kennedy, S., and Fire, A.Z. (2010). Distinct phases of siRNA synthesis in an endogenous RNAi pathway in *C. elegans* soma. *Mol. Cell* 37, 679–689.

Ghildiyal, M., and Zamore, P.D. (2009). Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.* 10, 94–108.

Gu, W., Shirayama, M., Conte, D., Jr., Vasale, J., Batista, P.J., Claycomb, J.M., Moresco, J.J., Youngman, E.M., Keys, J., Stoltz, M.J., et al. (2009). Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the *C. elegans* germline. *Mol. Cell* 36, 231–244.

Gu, W., Lee, H.C., Chaves, D., Youngman, E.M., Pazour, G.J., Conte, D., Jr., and Mello, C.C. (2012). CapSeq and CIP-TAP identify Pol II start sites and reveal capped small RNAs as *C. elegans* piRNA precursors. *Cell* 151, 1488–1500.

Guang, S., Bochner, A.F., Pavelec, D.M., Burkhart, K.B., Harding, S., Lachowicz, J., and Kennedy, S. (2008). An Argonaute transports siRNAs from the cytoplasm to the nucleus. *Science* 321, 537–541.

Guang, S., Bochner, A.F., Burkhart, K.B., Burton, N., Pavelec, D.M., and Kennedy, S. (2010). Small regulatory RNAs inhibit RNA polymerase II during the elongation phase of transcription. *Nature* 465, 1097–1101.

Huntzinger, E., and Izaurralde, E. (2011). Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat. Rev. Genet.* 12, 99–110.

Ibrahim, F., Rymarquis, L.A., Kim, E.J., Becker, J., Balassa, E., Green, P.J., and Cerutti, H. (2010). Uridylation of mature miRNAs and siRNAs by the MUT68 nucleotidyltransferase promotes their degradation in *Chlamydomonas*. *Proc. Natl. Acad. Sci. USA* 107, 3906–3911.

Jain, C. (2012). Novel role for RNase PH in the degradation of structured RNA. *J. Bacteriol.* 194, 3883–3890.

Kao, C.C., Singh, P., and Ecker, D.J. (2001). De novo initiation of viral RNA-dependent RNA synthesis. *Virology* 287, 251–260.

Kennedy, S., Wang, D., and Ruvkun, G. (2004). A conserved siRNA-degrading RNase negatively regulates RNA interference in *C. elegans*. *Nature* 427, 645–649.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25.

Lee, S.R., Talsky, K.B., and Collins, K. (2009). A single RNA-dependent RNA polymerase assembles with mutually exclusive nucleotidyl transferase subunits to direct different pathways of small RNA biogenesis. *RNA* 15, 1363–1374.

Lee, H.C., Gu, W., Shirayama, M., Youngman, E., Conte, D., Jr., and Mello, C.C. (2012). *C. elegans* piRNAs mediate the genome-wide surveillance of germline transcripts. *Cell* 150, 78–87.

Lee, M., Kim, B., and Kim, V.N. (2014). Emerging roles of RNA modification: m(6)A and U-tail. *Cell* 158, 980–987.

Lim, J., Ha, M., Chang, H., Kwon, S.C., Simanshu, D.K., Patel, D.J., and Kim, V.N. (2014). Uridylation by TUT4 and TUT7 marks mRNA for degradation. *Cell* 159, 1365–1376.

Makeyev, E.V., and Bamford, D.H. (2002). Cellular RNA-dependent RNA polymerase involved in posttranscriptional gene silencing has two distinct activity modes. *Mol. Cell* 10, 1417–1427.

Matsushita, K., Takeuchi, O., Standley, D.M., Kumagai, Y., Kawagoe, T., Miyake, T., Satoh, T., Kato, H., Tsujimura, T., Nakamura, H., and Akira, S. (2009). Zc3h12a is an RNase essential for controlling immune responses by regulating mRNA decay. *Nature* 458, 1185–1190.

Olovnikov, I., Chan, K., Sachidanandam, R., Newman, D.K., and Aravin, A.A. (2013). Bacterial argonaute samples the transcriptome to identify foreign DNA. *Mol. Cell* 51, 594–605.

Pak, J., and Fire, A. (2007). Distinct populations of primary and secondary effectors during RNAi in *C. elegans*. *Science* 315, 241–244.

Pak, J., Maniar, J.M., Mello, C.C., and Fire, A. (2012). Protection from feed-forward amplification in an amplified RNAi mechanism. *Cell* 151, 885–899.

Pavelec, D.M., Lachowicz, J., Duchaine, T.F., Smith, H.E., and Kennedy, S. (2009). Requirement for the ERI/DICER complex in endogenous RNA interference and sperm development in *Caenorhabditis elegans*. *Genetics* 183, 1283–1295.

Phillips, C.M., McDonald, K.L., and Dernburg, A.F. (2009). Cytological analysis of meiosis in *Caenorhabditis elegans*. *Methods Mol. Biol.* 558, 171–195.

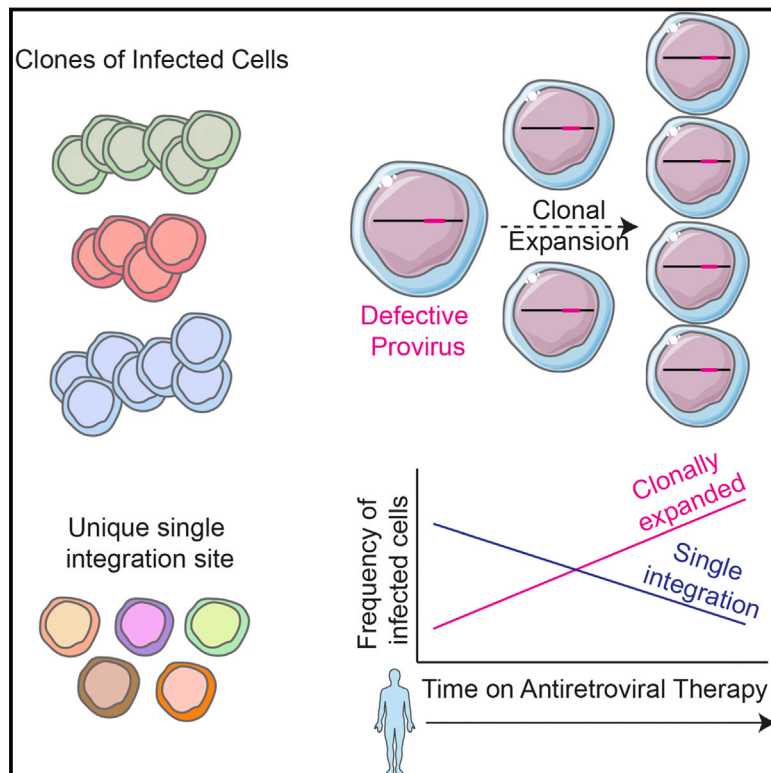
Phillips, C.M., Montgomery, T.A., Breen, P.C., and Ruvkun, G. (2012). MUT-16 promotes formation of perinuclear mutator foci required for RNA silencing in the *C. elegans* germline. *Genes Dev.* 26, 1433–1444.

Ruby, J.G., Jan, C., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H., and Bartel, D.P. (2006). Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* 127, 1193–1207.

- Seth, M., Shirayama, M., Gu, W., Ishidate, T., Conte, D., Jr., and Mello, C.C. (2013). The *C. elegans* CSR-1 argonaute pathway counteracts epigenetic silencing to promote germline gene expression. *Dev. Cell* 27, 656–663.
- Shen, B., and Goodman, H.M. (2004). Uridine addition after microRNA-directed cleavage. *Science* 306, 997.
- Shirayama, M., Seth, M., Lee, H.C., Gu, W., Ishidate, T., Conte, D., Jr., and Mello, C.C. (2012). piRNAs initiate an epigenetic memory of nonself RNA in the *C. elegans* germline. *Cell* 150, 65–77.
- Shirayama, M., Stanney, W., 3rd, Gu, W., Seth, M., and Mello, C.C. (2014). The Vasa Homolog RDE-12 engages target mRNA and multiple argonaute proteins to promote RNAi in *C. elegans*. *Curr. Biol.* 24, 845–851.
- Sijen, T., Fleenor, J., Simmer, F., Thijssen, K.L., Parrish, S., Timmons, L., Plasterk, R.H., and Fire, A. (2001). On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell* 107, 465–476.
- Sijen, T., Steiner, F.A., Thijssen, K.L., and Plasterk, R.H. (2007). Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. *Science* 315, 244–247.
- Smardon, A., Spoerke, J.M., Stacey, S.C., Klein, M.E., Mackin, N., and Maine, E.M. (2000). EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in *C. elegans*. *Curr. Biol.* 10, 169–178.
- Steiner, F.A., Okihara, K.L., Hoogstrate, S.W., Sijen, T., and Ketting, R.F. (2009). RDE-1 slicer activity is required only for passenger-strand cleavage during RNAi in *Caenorhabditis elegans*. *Nat. Struct. Mol. Biol.* 16, 207–211.
- Suzuki, H.I., Arase, M., Matsuyama, H., Choi, Y.L., Ueno, T., Mano, H., Sugimoto, K., and Miyazono, K. (2011). MCP1P1 ribonuclease antagonizes dicer and terminates microRNA biogenesis through precursor microRNA degradation. *Mol. Cell* 44, 424–436.
- Talsky, K.B., and Collins, K. (2010). Initiation by a eukaryotic RNA-dependent RNA polymerase requires looping of the template end and is influenced by the template-tailing activity of an associated uridylyltransferase. *J. Biol. Chem.* 285, 27614–27623.
- Thivierge, C., Makil, N., Flamand, M., Vasale, J.J., Mello, C.C., Wohlschlegel, J., Conte, D., Jr., and Duchaine, T.F. (2012). Tudor domain ERI-5 tethers an RNA-dependent RNA polymerase to DCR-1 to potentiate endo-RNAi. *Nat. Struct. Mol. Biol.* 19, 90–97.
- Vasale, J.J., Gu, W., Thivierge, C., Batista, P.J., Claycomb, J.M., Youngman, E.M., Duchaine, T.F., Mello, C.C., and Conte, D., Jr. (2010). Sequential rounds of RNA-dependent RNA transcription drive endogenous small-RNA biogenesis in the ERGO-1/Argonaute pathway. *Proc. Natl. Acad. Sci. USA* 107, 3582–3587.
- Vastenhouw, N.L., Fischer, S.E., Robert, V.J., Thijssen, K.L., Fraser, A.G., Kamath, R.S., Ahringer, J., and Plasterk, R.H. (2003). A genome-wide screen identifies 27 genes involved in transposon silencing in *C. elegans*. *Curr. Biol.* 13, 1311–1316.
- Wolke, U., Jezuit, E.A., and Priess, J.R. (2007). Actin-dependent cytoplasmic streaming in *C. elegans* oogenesis. *Development* 134, 2227–2236.
- Xu, J., Peng, W., Sun, Y., Wang, X., Xu, Y., Li, X., Gao, G., and Rao, Z. (2012). Structural study of MCP1P1 N-terminal conserved domain reveals a PIN-like RNase. *Nucleic Acids Res.* 40, 6957–6965.
- Yang, H., Zhang, Y., Vallandingham, J., Li, H., Florens, L., and Mak, H.Y. (2012). The RDE-10/RDE-11 complex triggers RNAi-induced mRNA degradation by association with target mRNA in *C. elegans*. *Genes Dev.* 26, 846–856.
- Yang, H., Vallandingham, J., Shiu, P., Li, H., Hunter, C.P., and Mak, H.Y. (2014). The DEAD box helicase RDE-12 promotes amplification of RNAi in cytoplasmic foci in *C. elegans*. *Curr. Biol.* 24, 832–838.
- Yigit, E., Batista, P.J., Bei, Y., Pang, K.M., Chen, C.C., Tolia, N.H., Joshua-Tor, L., Mitani, S., Simard, M.J., and Mello, C.C. (2006). Analysis of the *C. elegans* Argonaute family reveals that distinct Argonautes act sequentially during RNAi. *Cell* 127, 747–757.
- Zhang, C., Montgomery, T.A., Gabel, H.W., Fischer, S.E., Phillips, C.M., Fahlgren, N., Sullivan, C.M., Carrington, J.C., and Ruvkun, G. (2011). *mut-16* and other mutator class genes modulate 22G and 26G siRNA pathways in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* 108, 1201–1208.

HIV-1 Integration Landscape during Latent and Active Infection

Graphical Abstract



Authors

Lillian B. Cohn, Israel T. Silva, ...,
Mila Jankovic, Michel C. Nussenzweig

Correspondence

nussen@rockefeller.edu

In Brief

HIV-1-infected CD4⁺ T cells that undergo clonal expansion are able to proliferate because their proviruses are defective. Conversely, the replication-competent reservoir is likely found in the subset of CD4⁺ T cells that carry unique integrations and remain quiescent.

Highlights

- Integration sequencing identifies clonally expanded and single HIV-1 integrations in human subjects
- Large clonal families of HIV-1+ cells are likely not part of the latent reservoir
- HIV-1 integrates near or into a 30 bp INT-motif found in *Alu* repeats



HIV-1 Integration Landscape during Latent and Active Infection

Lillian B. Cohn,¹ Israel T. Silva,^{1,2} Thiago Y. Oliveira,¹ Rafael A. Rosales,³ Erica H. Parrish,⁴ Gerald H. Learn,⁴ Beatrice H. Hahn,⁴ Julie L. Czartoski,⁵ M. Juliana McElrath,⁵ Clara Lehmann,^{6,7} Florian Klein,¹ Marina Caskey,¹ Bruce D. Walker,^{8,9} Janet D. Siliciano,¹⁰ Robert F. Siliciano,^{9,10} Mila Jankovic,¹ and Michel C. Nussenzweig^{1,9,*}

¹Laboratory of Molecular Immunology, The Rockefeller University, New York, NY 10065, USA

²National Institute of Science and Technology in Stem Cell and Cell Therapy and Center for Cell Based Therapy, Rua Catão Roxo, 2501, Ribeirão Preto CEP 14051-140, Brazil

³Departamento de Computação e Matemática, Universidade de São Paulo. Av. Bandeirantes, 3900, Ribeirão Preto CEP 14049-901, Brazil

⁴Departments of Medicine and Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

⁵Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA

⁶Department I of Internal Medicine, University Hospital of Cologne, 50924 Cologne, Germany

⁷German Centre for Infection Research, partner site Bonn-Cologne, 50924 Cologne, Germany

⁸Ragon Institute of MGH, Cambridge, MA 02139, USA

⁹Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA

¹⁰Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

*Correspondence: nussen@rockefeller.edu

<http://dx.doi.org/10.1016/j.cell.2015.01.020>

SUMMARY

The barrier to curing HIV-1 is thought to reside primarily in CD4⁺ T cells containing silent proviruses. To characterize these latently infected cells, we studied the integration profile of HIV-1 in viremic progressors, individuals receiving antiretroviral therapy, and viremic controllers. Clonally expanded T cells represented the majority of all integrations and increased during therapy. However, none of the 75 expanded T cell clones assayed contained intact virus. In contrast, the cells bearing single integration events decreased in frequency over time on therapy, and the surviving cells were enriched for HIV-1 integration in silent regions of the genome. Finally, there was a strong preference for integration into, or in close proximity to, *Alu* repeats, which were also enriched in local hotspots for integration. The data indicate that dividing clonally expanded T cells contain defective proviruses and that the replication-competent reservoir is primarily found in CD4⁺ T cells that remain relatively quiescent.

INTRODUCTION

Despite effective therapy, HIV-1 can persist in a latent state as an integrated provirus in resting memory CD4⁺ T cells (Chun et al., 1997; Finzi et al., 1997; Wong et al., 1997). The latent reservoir is established very early during infection, (Chun et al., 1998), and because of its long half-life of 44 months (Finzi et al., 1999), it is the major barrier to curing HIV-1 infection (Siliciano and Greene, 2011).

The HIV-1 latent reservoir has been difficult to define, in part because reactivation of latent viruses is difficult to induce and

to measure. Viral outgrowth assays underestimate the size of the reservoir, while direct measurements of integrated HIV-1 DNA overestimate the reservoir because a large fraction of the integrated viruses are defective (Ho et al., 2013). Although the latent reservoir remains to be completely defined, establishing the reservoir requires intact retroviral integration into the genome and subsequent transcriptional silencing (Siliciano and Greene, 2011). Whether or not the genomic location of the integration impacts on latency is debated (Jordan et al., 2003; Jordan et al., 2001; Sherrill-Mix et al., 2013). However, HIV integration into the genome is known to favor the introns of expressed genes (Han et al., 2004), some of which, like *BACH2* and *MKL2*, carry multiple independent HIV-1 integrations in different individuals and are considered hotspots for integration (Ikeda et al., 2007; Maldarelli et al., 2014; Wagner et al., 2014). However, there is currently no precise understanding of the nature of these hotspots or why they are targeted by HIV-1.

Viremia rebounds from the latent reservoir after interruption of long-term treatment with combination anti-retroviral therapy (cART). When it does, it appears to involve an increasing proportion of monotypic HIV-1 sequences, suggesting the proliferation of latently infected cells (Wagner et al., 2013). Based on this observation and the finding that a subset of cells bearing integrated HIV-1 undergoes clonal expansion in patients receiving suppressive anti-retroviral therapy, it has been proposed that the clonally expanded cells play a critical role in maintaining the reservoir (Maldarelli et al., 2014; Wagner et al., 2014).

To obtain additional insights into the regions of the genome that are favored by HIV-1 for integration and the role of clonal expansion in maintaining the reservoir, we developed a single cell method to identify a large number of HIV-1 integration sites from treated and untreated individuals, including “viremic controllers” who spontaneously maintain viral loads of <2,000 RNA copies/ml and “typical progressors” who display viral loads >2,000 RNA copies/ml.

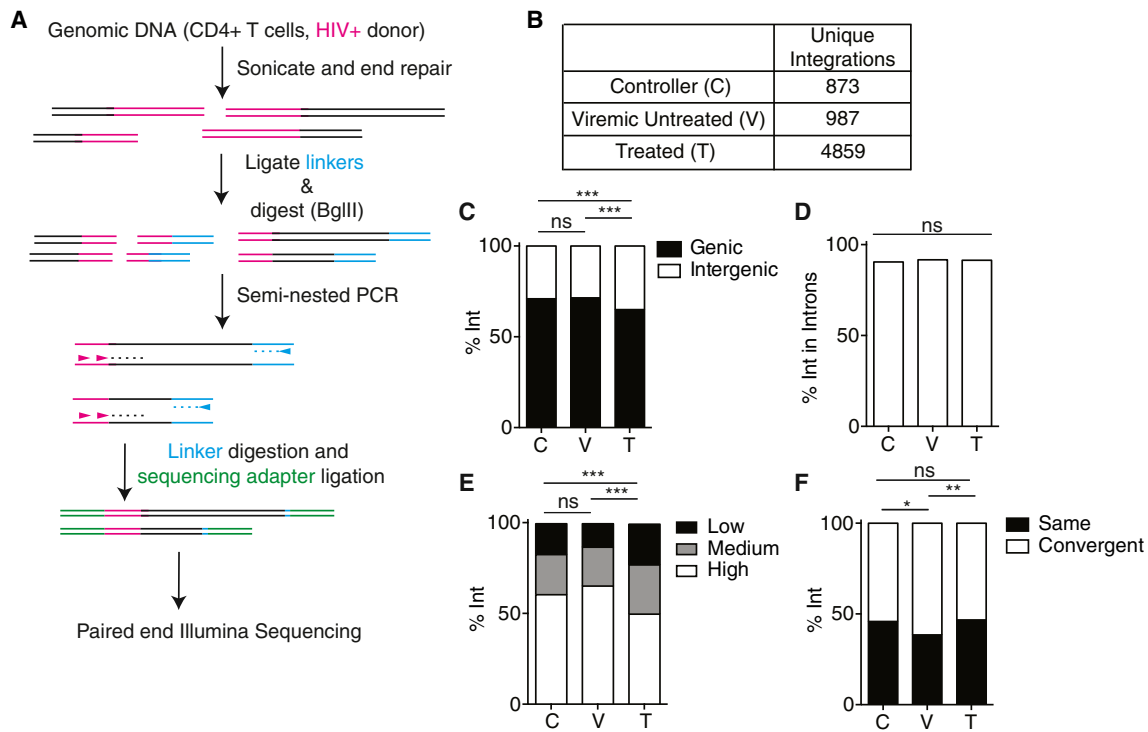


Figure 1. HIV-1 Integration Libraries

(A) Diagram of integration library construction.

(B) Table of unique integrations identified in viremic controllers (C), viremic untreated progressors (V), and treated progressors (T).

(C) Proportion of integrations (Int) that are in genic or intergenic regions.

(D) Proportion of genic integrations located in introns.

(E) Proportion of integrations in genes with high, medium, or low expression. p values refer to proportion of integrations in highly expressed genes.

(F) Transcriptional orientation of integrated HIV-1 relative to host gene. ns, not significant.

*p < 0.05; **p < 0.01; ***p < 0.0001 using two-proportion z test. See also Figure S1.

RESULTS

Integration Library Construction

Twenty-four integration libraries were constructed from CD4⁺ T cells from 13 individuals: three provided longitudinal samples before and after (0.1–7.2 years) initiation of therapy; four were untreated; two were treated; and four were viremic controllers (Table S1). Patients were grouped into three categories based on viral loads and therapy: (1) viremic progressors were untreated individuals with viral loads higher than 2,000 viral RNA copies/ml of plasma; (2) progressors were treated individuals whose initial viral loads were higher than 2,000 viral RNA copies/ml before therapy; (3) controllers were individuals who maintain low viral loads spontaneously in the absence of therapy (<2,000 viral RNA copies/ml). The frequency of latently infected, resting CD4⁺ T cells in our patients was similar to that reported by others as measured by quantitative viral outgrowth assay (Table S1 and Laird et al., 2013).

Libraries were produced from genomic DNA by a modification of the translocation-capture sequencing method that we refer to in this paper as integration sequencing (Figure 1A) (Janovitz et al., 2013; Klein et al., 2011). Virus integration sites were recovered by semi-nested ligation-mediated PCR from fragmented DNA using primers specific to the HIV-1 3' LTR (Table S2).

PCR products were subjected to high-throughput paired-end sequencing, and reads were aligned to the human genome. Since sonication is random, it produces unique linker ligation points that identify the specific integration events in each infected CD4⁺ T cell, which allows both single-cell resolution and identification of expanded clones of cells with identical integrations (Berry et al., 2012 and Figure 1A). Thus, integration sequencing can enumerate both the number of integration sites and the number of infected cells.

A total of 6,719 unique virus integration sites were determined (Table S3): 873 unique integrations in viremic controllers; 987 integrations in untreated progressors; and 4,859 integrations in treated progressors (Figure 1B).

Integrations Are Enriched in Introns of Highly Expressed Genes

We analyzed the genomic location of the integration sites obtained from viremic controllers and untreated and treated progressors and compared our results to published data obtained from HIV-1-infected individuals (Han et al., 2004; Ho et al., 2013; Ikeda et al., 2007; Schröder et al., 2002). In agreement with the work of others, the majority of integration sites in each group are genic (Figure 1C and Figure S1A). Moreover,

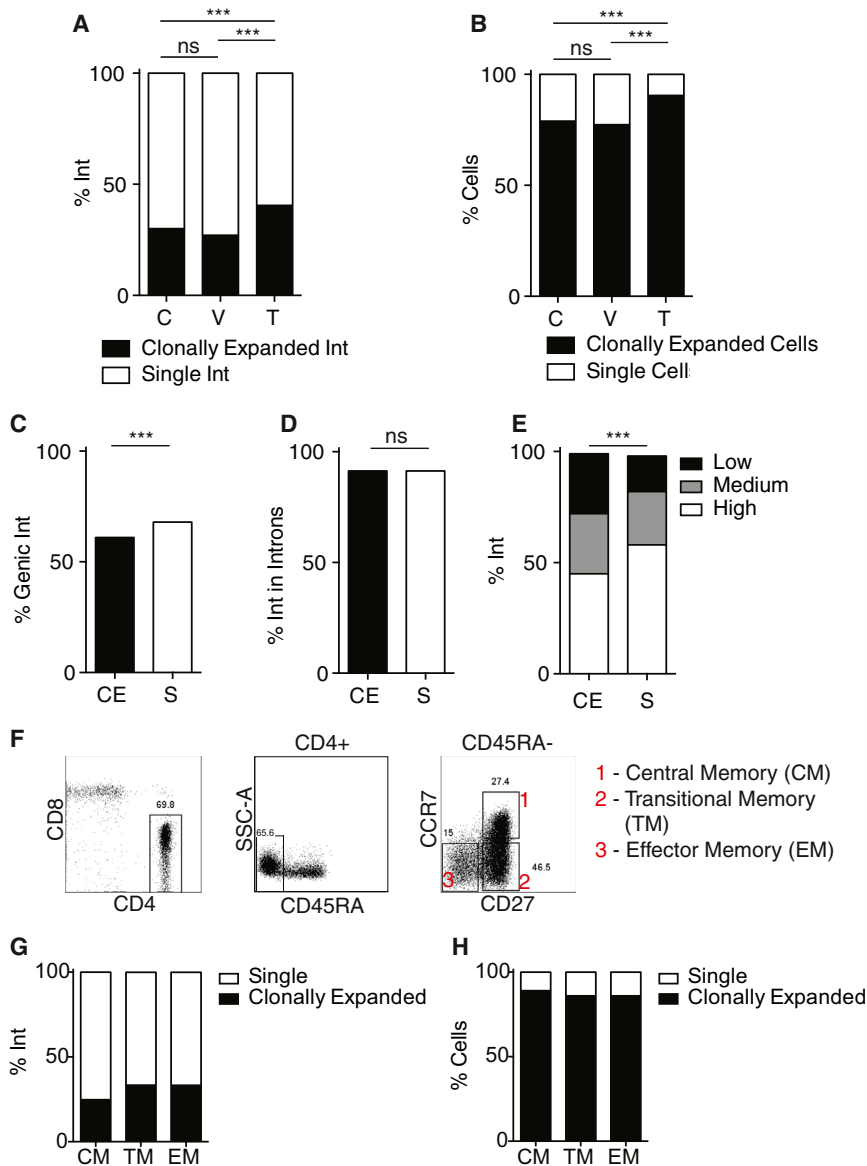


Figure 2. Identification of Clonally Expanded Cells Bearing Integrated HIV-1

(A) Proportion of viral integrations (Int) that are clonally expanded, as identified by the same integration site with multiple shears in controllers (C) or viremic (V) or treated progressors (T). (B) Proportion of infected cells deriving from clonal expansion.

(C) Proportion of clonally expanded (CE) and single (S) viral integrations in genic or intergenic regions. (D) Proportion of clonally expanded and single viral integrations in introns.

(E) Proportion of clonally expanded or single viral integrations in genes with high, medium, or low expression. p values refer to proportion of integrations in highly expressed genes.

(F) Flow cytometry sorting strategy to identify CD4⁺ T cell subsets. CM, TM, and EM cell subsets were identified based on their CD45RA, CCR7, and CD27 expression. Shown is one representative sort.

(G) Proportion of viral integrations (Int) that are clonally expanded, as identified by the same integration site with multiple shears in sorted CD4⁺ T cell subsets from patient 9.

(H) Proportion of infected cells deriving from clonal expansion in sorted CD4⁺ T cell subsets from patient 9.

ns, not significant. ***p < 0.0001 using two-proportion z test. See also Figure S2.

controllers and treated progressors were not significantly different from each other in terms of the level of expression of the genes at the sites of integration (Figure 1E). Thus, therapy is associated with a relative decrease in the number of cells with viral integrations in highly expressed genes.

Identification of Clonally Expanded Cells Containing Integrated HIV-1

Since we shear DNA ends randomly to produce our libraries and by paired-end

sequencing can determine the precise site of both the integration and sheared end, we infer that identical integrations with unique sheared ends arise from clones of expanded cells (Figure 1A). Integrations can therefore be classified as clonally expanded (i.e., identical integrations with distinct sheared ends, deriving from the clonal expansion of an original unique, single integration event) or single integrations (i.e., unique integration site with a single sheared end).

Clonally expanded viral integrations were present in all individuals irrespective of therapy or viremia (Table S3). However, the proportion of clonally expanded viral integrations is significantly lower in viremic controllers (30%) and viremic progressors (27%) than in treated progressors (40%) (p < 0.0001 and p < 0.0001, Figure 2A and Figure S2A). Although the size of individual clones varied from 2 to 295 cells (Figure S2B), the relative increase in clonally expanded integrations during therapy consistently

integrations are found more frequently in the introns of highly expressed genes, and there is a slight bias for viral orientation that leads to convergent transcription (Figures 1D, 1E, and 1F and Figures S1B–S1E) (Mitchell et al., 2004; Schröder et al., 2002). Thus, the general features of integrations defined by integration sequencing are similar to those obtained by others.

Although the differences between groups were small in magnitude, they were significant in that treated progressors had a smaller proportion of integrations in genic regions (p < 0.0001 and p < 0.0001, respectively) and in highly expressed genes (p < 0.0001 and p < 0.0001, respectively) when compared to viremic controllers and untreated progressors (Figures 1C and 1E). Conversely, the proportion of viral integrations in genes expressed at lower levels was increased in treated progressors compared to viremic controllers and untreated progressors (p = 0.002 and p < 0.0001, respectively). Viremic

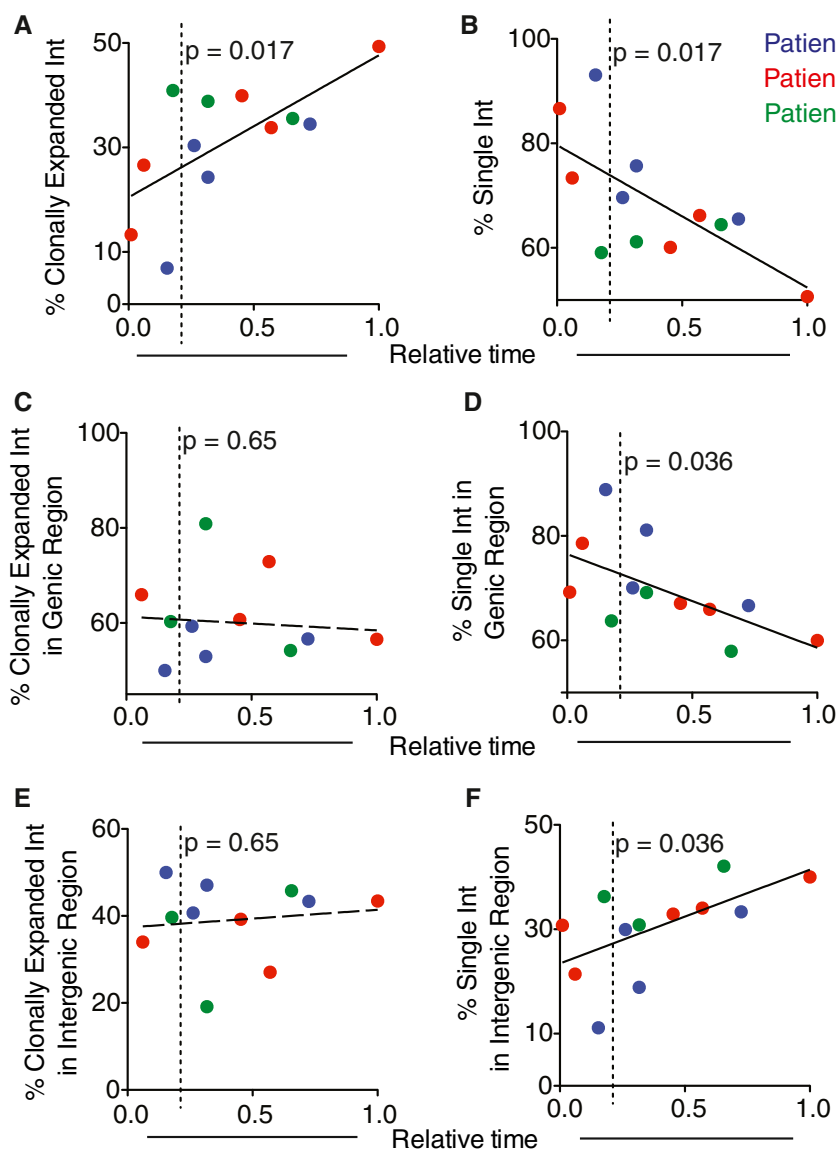


Figure 3. Clonally Expanded Viral Integrations Increase and Single Integrations Decrease during Therapy

Graphs show data from patients 1 (blue), 2 (red), and 3 (green) from longitudinal time points (Table S1). Time was normalized from 0 to 1 (727 days pre-therapy to 2617 days post-therapy). Dotted line at $t = 0.21$ marks therapy initiation. Trendline was determined by linear regression model. Solid lines indicate significant change in proportion of events; dashed lines indicate insignificant change in proportion of events.

(A) Proportion of clonally expanded viral integrations (Int).

(B) Proportion of single viral integrations.

(C) Proportion of genic clonally expanded viral integrations.

(D) Proportion of genic single viral integrations.

(E) Proportion of intergenic clonally expanded viral integrations.

(F) Proportion of intergenic single viral integrations. See also Figure S3.

more likely to be found in highly expressed genes than clonal integrations (Figure 2E, $p < 0.0001$, and Figure S2F). Thus, cells harboring viral integrations in intergenic regions and genes that are expressed at lower levels are more likely to be clonally expanded.

Large Expanded Clones Are Found in Memory Cells

Central memory cells are thought to be the major source of the HIV-1 reservoir (Chomont et al., 2009). To investigate the nature of the cells that comprise the expanded clones, we performed virus integration sequencing on genomic DNA from sorted central, transitional, and effector memory CD4⁺ T cells (Figure 2F). In both individuals studied, all three subsets of CD4⁺ T cells contained expanded

translated into an increase in the number of infected cells that derive from expanded clones (Figure 2B). The percentage of cells containing clonally expanded HIV-1 integrations was similar in untreated progressors (78%) and controllers (79%), but it was significantly increased in treated progressors (90%) ($p < 0.0001$ and $p < 0.0001$, Figure 2B and Figure S2C). Thus, therapy is associated with an increase in the frequency of clonal HIV-1 integrations and infected clonally expanded cells.

To determine whether the position of viral integration in the genome correlates with clonal expansion, we compared the location of genomic clonally expanded to single integrations. Both types of integrations favored genes and their introns (Figures 2C and 2D and Figures S2D and S2E). However, the proportion of clonally expanded integrations in intergenic regions was greater than that of single integrations (Figure 2C, $p < 0.0001$). Moreover, of the integrations in genes, single integrations were

clones (Figures 2G and 2H and Figures S2G and S2H). Thus, central, transitional, and effector memory T cells, all of which have undergone antigen-stimulated cell division, harbor the expanded clones of HIV-1 integrants.

Clonally Expanded Integrations Increase after Therapy

The proportion of clonally expanded viral integrations is increased in treated progressors (Figure 2A and Wagner et al., 2014). To further examine the effect of therapy on clonal expansion, we analyzed longitudinal samples from three typical progressors before and during therapy (Table S1). We found an increase in the number of clonally expanded integrations throughout the treatment period of up to 7.2 years in two of the three patients (Figure 3A, $p = 0.017$, and Figure S3A) as well as an increase in the number of cells that contained clonally expanded viral integrations (Figure S3B). Correspondingly, there

was also an overall decrease over time in single integrations ($p = 0.017$), with a half-life of 127 months assuming a non-linear regression model for one-phase decay (Figure 3B). Thus, our data suggest that the numbers of single integrations decay very slowly over time, while clonally expanded integrations increase with time on cART.

The increase in the number of clonal integrations during cART did not favor genic or intergenic regions ($p = 0.65$), indicating that this effect is independent of the location of the integration in the genome (Figures 3C and 3E and Figure S3C). In contrast, single integrations decrease significantly in genic regions (Figure 3D, $p = 0.036$, and Figure S3D) and increase proportionally in intergenic regions (Figure 3F, $p = 0.036$). Thus, the fate of cells harboring single viral integrations in cART treated progressors differs from clonal integration. Moreover, the fate of single integrations is dependent on their location in the genome, whereas the clonal integrations are not. These results suggest that cells bearing genic single integrations are selected against during therapy and that clonal expansion is not.

Clonally Expanded Integrations in the Same Genes in Multiple Patients

In the three progressors who provided longitudinal samples, ~5% of the clonal integrations persisted through successive time points without selection for genic or intergenic regions compared to all clonal integrations (Figures 4A and 4B). Furthermore, of the genic integrations that persisted, there was also no selection for or against those in highly expressed genes (Figure 4C). Thus, the persistent clonal integrations are indistinguishable from the larger pool of clonally expanded viral integrations in terms of their position in the genome.

To determine whether specific genes or groups of genes are permissive for clonal expansion, we looked for overlap in genic integration sites between samples (Figures 4D–4F). Despite a higher number of single integrations, there was much greater overlap of the genes that harbor clonally expanded integrations between individuals irrespective of treatment or level of viremic control ($p < 0.0001$) (Figures 4D–4F). On average, there is 13% and 3% overlap between genes harboring clonally expanded and single viral integrations, respectively. The genes containing clonally expanded viral integrations in multiple patients are expressed at lower levels than genes containing overlapping single viral integrations (Figure 4G). Taken together, these results suggest that cells that carry integrations in highly expressed genes tolerate clonal expansion less well than cells with integrations in genes with lower levels of expression.

Since clonal integrations have been associated with genes involved in malignant transformation (Wagner et al., 2014), we examined our entire dataset for enrichment of integrations in cancer-associated genes ($n = 743$ cancer-associated genes [Vogelstein et al., 2013; Zhao et al., 2013]). Although there was an overall enrichment for integrations in cancer genes ($329/4,410 = 7.5\%$) compared to all genes in the human genome ($743/25,660 = 2.8\%$) ($p < 0.0001$), this preference does not seem to be significant because it is similar to the overall preference for integration into highly expressed genes (Figure S4). Furthermore, we observed no overrepresentation of single, clonal, or persistent integrations in cancer genes (Figure 4H).

Importantly, a significant decrease in integrations in cancer-related genes was observed in longitudinal samples (Figure 4I), suggesting that these are selected against with therapy.

Expanded Clones Contain Defective Viruses

Our method of integration sequencing captures the end of the 3' LTR and identifies the genomic site of viral integration. To determine whether the viruses found in expanded clones are intact, we used nested integration site-specific PCR primers that were anchored in the host genome to amplify the 5' LTRs of 75 expanded clones from eight individuals (Figure 5A and Table S2). The clones selected for PCR verification varied in size from 5 to 200 out of $0.3\text{--}2 \times 10^6$ CD4 T cells. Of the 75 sequences obtained, 24 showed fragmented 5' LTRs flanked by the correct genomic site, and an additional 44 of the proviruses did not have a recoverable 5' end (Figure 5B). The remaining eight viruses with intact 5' LTRs were amplified in limiting dilution conditions using integration site-specific primers and HIV-1 primers (Figure 5C). Three of the eight viruses could not be amplified; four had large deletions in Env, one had a frameshift mutation in *pol*, and one had undergone APOBEC3G-mediated hypermutation to produce a premature stop codon in *env* (Figure 5D and Data S1). Thus, we were unable to find a single intact integrated provirus among 75 expanded clones.

Hotspots for Virus Integration

Overlap between integrations in the genes of different patients suggests the existence of hotspots for HIV-1 integration. A number of individual genes have been identified as preferential sites for HIV-1 integration, including *BACH2*, *MKL2*, *DMNT1*, *MDC1*, and *STAT5B* (Ikeda et al., 2007; Maldarelli et al., 2014; Wagner et al., 2014). To identify hotspots for HIV-1 integration genome-wide, we subjected our dataset to hot_scan analysis (Silva et al., 2014), which defines hotspots by identifying regions of local enrichment using scan statistics. This analysis identified 55, 85, and 247 hotspots for controllers, viremic, and treated progressors, respectively (Figure 6A). For example, the intron between exons 5 and 6 in *MKL2* is a hotspot for integration in patient 11, contains an expanded clonal family in patient 10, and was also identified as a site of enrichment for integration by others (Maldarelli et al., 2014) (Figure 6B).

To validate our *in silico* analysis and to further characterize the *MKL2* hotspot, we sequenced the *gag* gene from proviruses integrated into *MKL2* by amplification with nested genomic primers specific for *MKL2* and HIV-1 *gag*. Sequences obtained from patient 10, who showed only one expanded clone, are very closely related to each other, which is consistent with a single clonally expanded integration (Figure 6C). In contrast, sequences obtained from patient 11 are far more diverse, suggesting that there were several different viral integrations in the *MKL2* hotspot (Figure 6C). We conclude that the hotspots defined by hot_scan represent multiple distinct integration events in close proximity.

Viremic progressors had the highest proportion of integration events in hotspots, indicating that, in the case of high-level viremia, there are specific genomic locations that favor integration (Figure 6D). Although the majority of all integrations fall outside of hotspots (Figure 6D), hotspots resemble other integrations in that

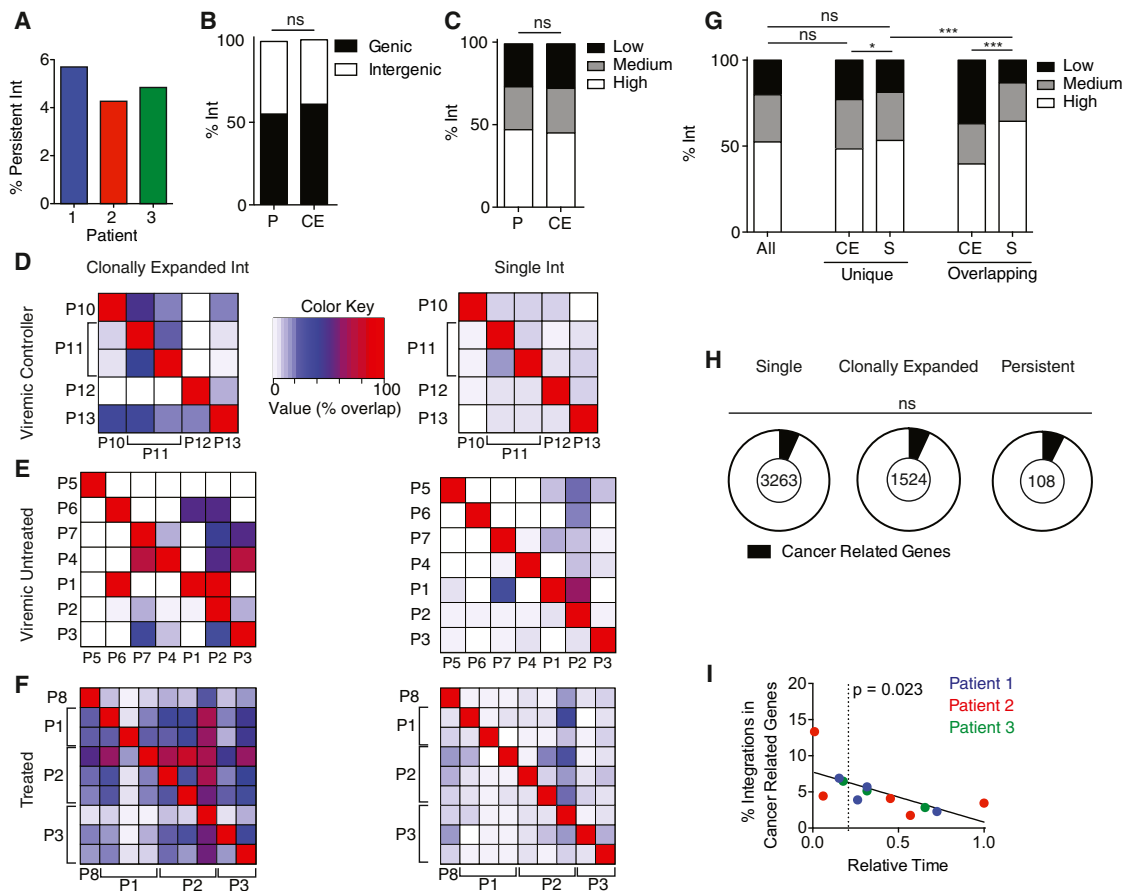


Figure 4. Integrations in Genes Permissive for Clonal Expansion Occur in Multiple Patients

(A) Percent viral integrations present in more than one time point (persistent integrations) in patients 1, 2, and 3 (Table S1).

(B) Comparison of persistent (P) and clonally expanded (CE) viral integrations in genic or intergenic regions.

(C) Proportion of persistent and clonally expanded viral integrations in genes with high, medium, or low expression. p values refer to proportion of integrations in highly expressed genes.

(D–F) Heatmap showing overlap between samples of genes containing clonally expanded or single viral integrations between samples. Patients are indicated by P1–13. Multiple samples from one individual are marked by a bracket. The amount of overlap is denoted by color (see legend); red, 100% overlap.

(G) Genes containing single or clonally expanded viral integrations were analyzed for their presence in multiple patients. Genes with integrations in more than one individual were classified as “overlapping”; genes with integrations in only one individual were classified as “unique.” Shown is the proportion of single and clonally expanded unique and overlapping viral integrations in genes with high, medium, or low expression. p values refer to proportion of integrations in highly expressed genes.

(H) Genes with integrations were analyzed for their association with cancer. Proportions of cancer-associated genes are shown for single, clonally expanded, and persistent viral integrations. The number indicates the total number of genes from each category.

(I) Graph shows proportion of integrations in cancer-related genes from patients 1 (blue), 2 (red), and 3 (green) from longitudinal time points (Table S1). Time was normalized from 0 to 1 (727 days pre-therapy to 2617 days post-therapy). Dotted line at $t = 0.21$ marks therapy initiation. Trendline was determined by linear regression model and indicates significant change in proportion of events. $p = 0.023$.

ns, not significant. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.0001$ using two-proportion z test. See also Figure S4.

they are preferentially found within genes with a preponderance of these in introns (Figures 6E and 6F). In all cases, hotspots are enriched in highly expressed genes, and consistent with the overall decrease in viral integrations in highly expressed genes during therapy, the proportion of hotspots in these genes also decreases (Figures 6G and 1E). Thus, the general characteristics of hotspots are similar to features of all integrations.

To determine whether there is a relationship between hotspots and clonally expanded viral integrations, we enumerated single and clonally expanded integrations in hotspots (Figure 6H).

Only a small fraction (11%–18%) of all single integrations were found in hotspots with untreated viremic progressors showing the highest level (Figure 6H). In contrast, there was a much higher proportion of clonal integrations in hotspots (30%–46%), with the lowest proportion in treated progressors (Figure 6H). This observation is consistent with the finding that there is a greater degree of overlap in genes that harbor clonally expanded rather than single integrations (Figures 4D–4F) and that clonally expanded integrations are more likely to be hotspots than single integrations (Figure 6H, $p < 0.0001$).

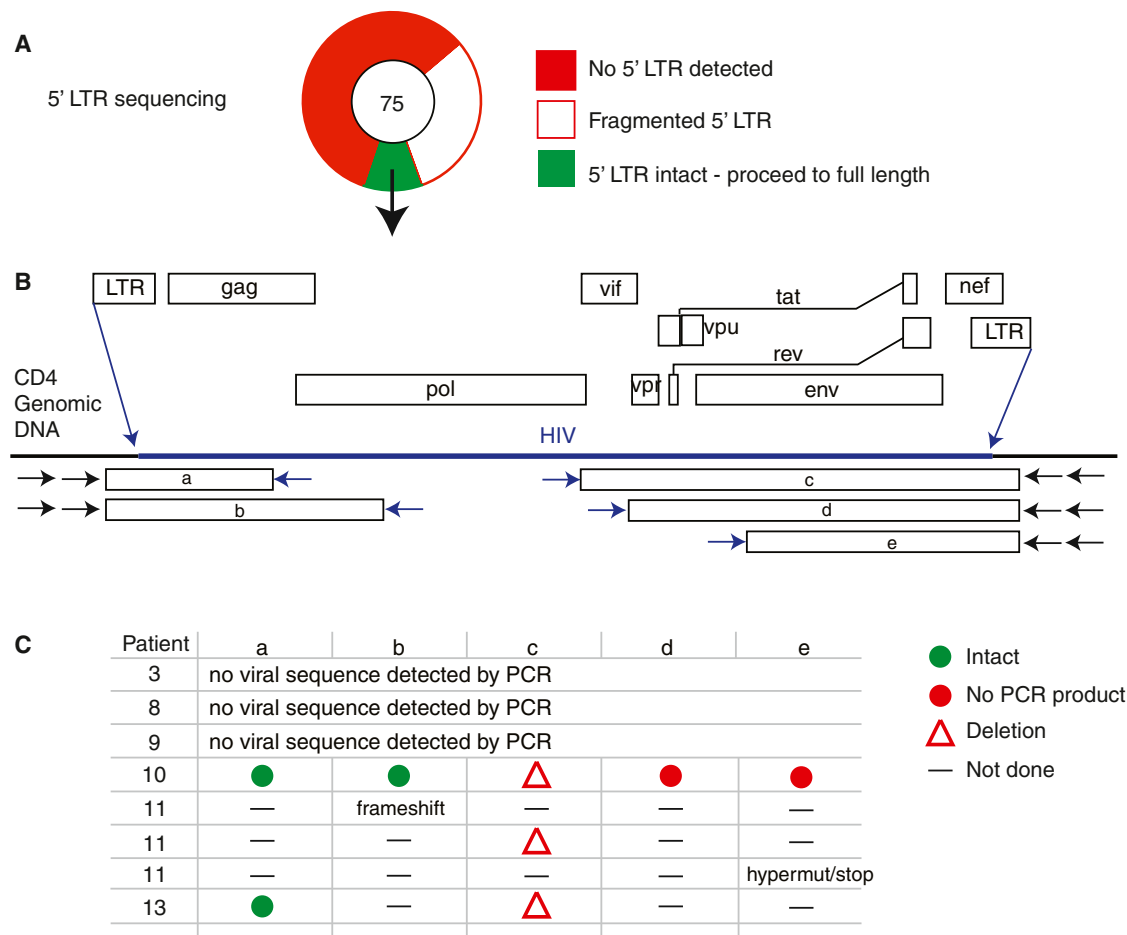


Figure 5. Large Expanded Clones Are Defective

(A) Sequence analysis of 5' LTRs in clonally expanded integrations. Of 75 different clonally expanded integrations from 8 individuals, 24 showed fragmented 5' LTRs, 44 didn't have a recoverable 5' LTR, and 8 contained intact 5' LTRs.

(B) Strategy for HIV-1 sequencing. Eight proviruses were analyzed for intact viral sequence. Nested genomic primers and internal HIV primers were used in a PCR walking strategy to amplify fragments a–e from specific clonally expanded integrations. PCR products were sequenced directly.

(C) Summary of HIV-1 sequencing from large expanded clones. Sequences were aligned to HXB2 and examined for the presence of large internal deletions. Intact sequences were analyzed for G→A hypermutation by Los Alamos Hypermut algorithm (Rose and Korber, 2000). Non-hypermutated products were analyzed for intact reading frames and frameshift mutations by Los Alamos HIVQC.

Green dot, intact, non-hypermutated sequence; red dot, no PCR product recovered; red triangle, sequence with internal deletion; —, not done. See also Data S1.

Viral Integration Enriched in Sites Containing a DNA Sequence Motif

To determine whether there are specific genomic features associated with sites of viral integration and hotspots, we examined 100 base pairs (bp) centered on all integration sites for the presence of a consensus sequence (Bailey and Elkan, 1994). We found 7% of all integrations within 100 bp of a highly conserved 30 bp motif (INT-motif) (Figure 7A). The majority of the integrations identified in this analysis were single integration events, with the ratio of single-to-clonal integrations being significantly different from the expected (Figure 7B, $p < 0.0001$). When HIV-1 integrates directly into the INT-motif, the 5' end of the motif is recurrently found 20 bp from the site of viral integration (Figure 7C). The INT-motif is asymmetrically distributed in *Alu* repeats, and its position coincides with a peak of viral integration (Figure 7D). Furthermore, there is a significant overall enrichment

of integrations inside *Alu* repeats (Figure 7E) and in close proximity to *Alu* repeats, irrespective of whether the integration is inside genes or in intergenic regions (Figure 7F). Thus, a preference for *Alu* is independent of a preference for integration in genes.

Previous studies have shown a preference for integration into *Alu* repeats, potentially because *Alu* repeats are enriched in gene-rich regions (Schröder et al., 2002). To further examine the relationship between *Alu* repeats and transcription of genes, we determined the distance between *Alu* repeats and the center of all genes. There was no positive correlation between the position of *Alu* and the level of transcription (Figure 7G). To determine whether the distance between integration and *Alu* repeats correlates with transcription, we measured the distance between the sites of integration and *Alu* repeats in all genes (Figure 7H). There was no significant

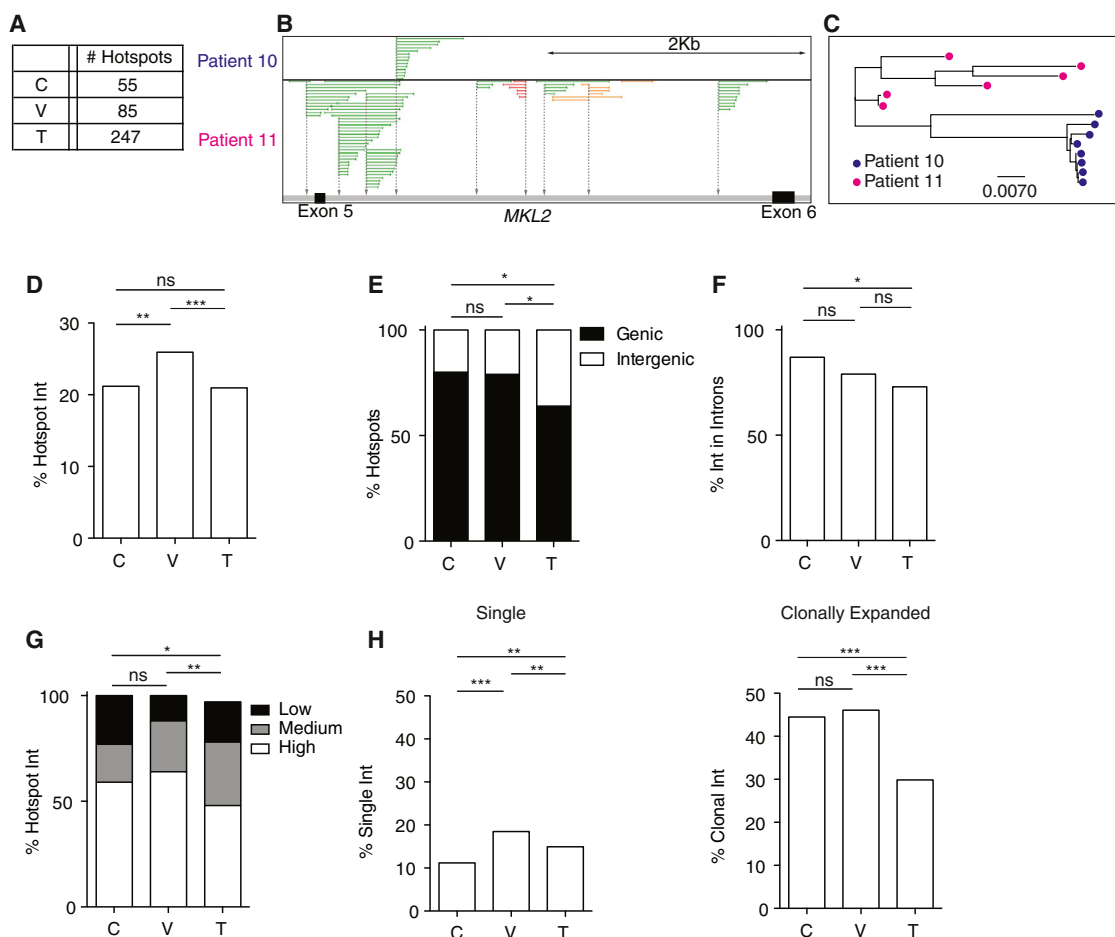


Figure 6. Identification of Hotspots for HIV-1 Integration

(A) Number of hotspots identified by hot-scan in viremic controllers (C) and viremic untreated (V) and treated progressors (T).

(B) Integrations in *MKL2* from patients 10 and 11. Gray vertical arrows indicate site of integrations. Colored horizontal lines show fragments of DNA spanning the point of integration through sheared end. Green, viruses integrated in the same orientation as gene; red, convergent orientation; orange, viruses integrated with both orientations.

(C) HIV-1 *gag* was amplified from integrated proviruses in *MKL2* from patients 10 and 11. PCR was performed using nested integration site-specific primers and HIV-1 *gag* primers. Sequences were clustered to assess DNA sequence similarity. The scale bar represents 0.007 substitutions per site.

(D) Proportion of virus integrations inside hotspots.

(E) Proportion of hotspots in genic and intergenic regions.

(F) Proportion of hotspots in introns.

(G) Proportion of hotspots in genes with high, medium, or low expression. p values refer to proportion of integrations in highly expressed genes.

(H) Percentage of total single and clonally expanded viral integrations inside of hotspots. Enrichment of clonally expanded viral integrations compared to single integrations is significant, $p < 0.0001$.

ns, not significant. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.0001$ using proportion test.

difference between integration distance to *Alu* repeats in highly expressed, silent, or trace level expressed genes. Therefore, the rate of transcription does not impact integration distance to *Alu* repeats, and integration at these sites must be independent of transcription.

Finally, the number of *Alu* repeats in a hotspot is directly correlated with the number of integration events in that hotspot (Figure 7I, $\rho = 0.86$). We conclude that HIV-1 has a preference for integration in close proximity to sites in the genome that are enriched in *Alu* repeats and that this preference is independent of the level of transcription.

DISCUSSION

CD4⁺ T cells that are actively infected with HIV-1 are rapidly eliminated during anti-retroviral therapy, but this form of treatment is relatively ineffective in selecting against latently infected CD4⁺ T cells, which have an estimated half-life of 44 months. Abolishing the latent reservoir is the current hurdle to finding a cure for HIV-1 infection. Although we have learned a great deal about the location of the latent compartment and its persistence during therapy, it has been difficult to uncover whether there are specific genomic features associated with latency (Siliciano and

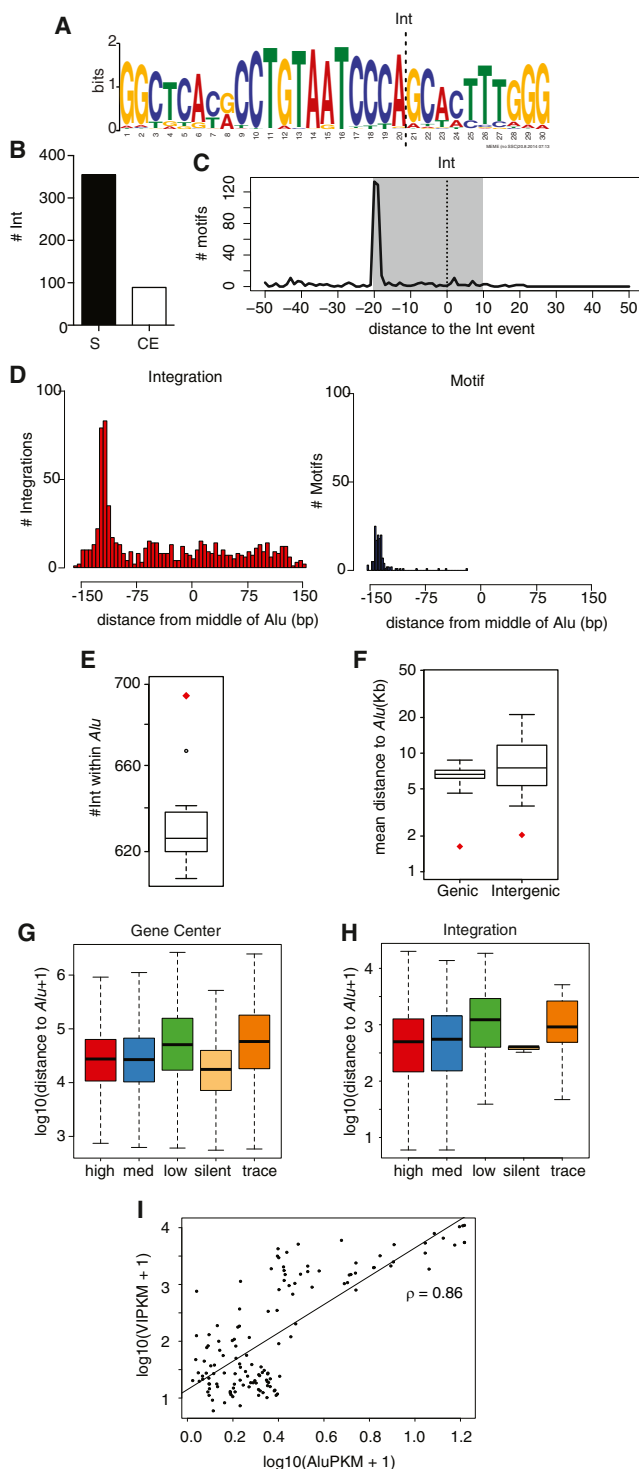


Figure 7. Consensus Motif for Viral Integration

(A) 30 bp sequence consensus motif (INT motif). 100 bp around all viral integration sites were analyzed for a consensus sequence by MEME (Bailey and Elkan, 1994). 444 integration sites were identified with the INT motif. E value: 6.4×10^{-4071} . The dotted line shows the preferred site of integration (see also C). (B) Number of single (S) and clonally expanded (CE) that were identified to contain INT motif within 100 bp of the integration site. $p < 0.0001$, using two-proportion z test.

(C) Conserved integration site within INT motif. Histogram maps the start site (5' end) of INT motif with respect to the integration site (dotted line). Peak shows that the majority of integration sites occur 20 bp from the 5' end of the motif start site. Shaded region represents the location of the INT motif relative to the majority of the integration sites.

(D) Location of integration preference and INT-motif inside *Alu* repeats is overlapping. (Left) Location of integration site *Alu* repeats were plotted relative to the midpoint of the repeat. (Right) The location of the start site of INT motifs within *Alu* repeats.

(E) Integrations are enriched inside *Alu* repeats. Total integrations identified inside *Alu* repeats were enumerated (red diamond) and compared to the expected value as defined by Monte Carlo simulation. The boxplot displays the variation of the number of random integrations identified inside *Alu* repeats by each iteration of the simulation. (F) Integrations are near *Alu* repeats in genes and intergenic regions. Average distance to the nearest *Alu* repeat for all integrations inside genes or intergenic regions was calculated (red diamond) and compared to the expected distance as defined by Monte Carlo simulation. The boxplot displays the variation of the distance of random integrations from *Alu* repeats in genes or intergenic regions by each iteration of the simulation. (G) Distance to *Alu* repeats from the center of highly, medium, low, trace, or silently expressed genes. (H) Distance to *Alu* repeats in highly, medium, low, trace, or silently expressed genes. (I) Positive correlation between *Alu* repeats and integrations inside hotspots. Graph shows number of *Alu* repeats (x axis) versus integrations in hotspots (y axis). Hotspots not containing *Alu* repeats were removed from this analysis. The scatter plot shows the linear relationship between the number of INT motifs and integrations inside hotspots (Pearson's correlation, $\rho = 0.86$).

experiments in cell lines showing that level of HIV-1 transcription is dependent, in part, on the status of surrounding chromatin (Jordan et al., 2003; Jordan et al., 2001; Sherrill-Mix et al., 2013).

HIV-1 integration has been studied in multiple cell types, but large libraries of integrations sites in primary infected T cell have only recently become available (Maldarelli et al., 2014; Wagner et al., 2014). Integration sites obtained from in-vitro-infected cell lines and primary T cells are distinct (Brady et al., 2009; Sherrill-Mix et al., 2013). Nevertheless, common features of HIV-1 integration have been defined, including the observation that integration favors *Alu* repeats (Schröder et al., 2002). This association was thought to be dependent on the presence of these repeats in the introns of gene-rich regions and not on a particular sequence feature (Schröder et al., 2002). However, we observed that integration preference into highly transcribed genes and into *Alu* repeats seem to be independently important, and furthermore, integrations are enriched near *Alu* repeats both in genic and intergenic regions. One possible explanation for preference for *Alu* seems to be the presence of an INT motif. TG-(N)₅₋₇-CA sequence has been associated with sites of HIV-1 integration, but an integration consensus has not been defined (Brady et al., 2009; Holman and Coffin, 2005; Lewinski et al., 2006; Serrao et al., 2014; Wang et al., 2007; Wu et al., 2005). We found a 30 bp INT motif within 100 bp of 7% of all integrations, the large majority of which are single events. As expected, the HIV-1 INT motif contains a signature TG-(N)₅₋₇-CA and can form a hairpin structure, anchored on 5' NTG-3', 5' CAN-3'. This motif is frequently found at the 3' end of *Alu*, where it coincides with a peak of viral integration events, and viruses integrated directly in this motif show a dramatic specificity for insertion site. The asymmetric peak and specificity of the integration site are remarkable. Nevertheless, we are likely underestimating the frequency of integrations within *Alu* because we can only map unique reads.

The observation that HIV-1 prefers to integrate in the neighborhood of *Alu* repeats is consistent with the finding that different individuals have been reported to have multiple integrations in selected genes (Ikeda et al., 2007; Maldarelli et al., 2014; Schröder et al., 2002; Wagner et al., 2014). Our experiments define a group of overlapping hotspots for integration that share many of the features of all HIV-1 integrations, including preference for introns of highly expressed genes and high density of *Alu* repeats. Viremic progressors showed the highest levels of hotspot integration, possibly because persistent integration leads to over-representation of these favored sites. Alternatively, these integrations might be positively selected by mechanisms that remain to be determined.

Individuals receiving cART show increasing numbers of cells with identical viral genomes by SGA, suggesting clonal expansion of a subset of cells bearing integrated proviruses (Buzon et al., 2014; Chomont et al., 2009; Wagner et al., 2013). Two independent groups have recently documented the long-term persistence of expanded clones of cells during therapy with cART (Maldarelli et al., 2014; Wagner et al., 2014). Our analysis confirms and extends these observations by showing that, when considered as a group, expanded clones are less likely to occur when the provirus is in a genic region, and clones that are associated with genes tend to be in genes that are expressed

at lower levels than single integrations. Thus, proviruses inserted into active regions of the genome, which would be more likely to support viral re-activation during T cell proliferation, are generally selected against during clonal expansion.

Why certain integration sites are permissive for clonal expansion is not known, but finding that expanded clones with integrations occur in cancer-related genes led to the suggestion that integration into genes that regulate cell division promotes proliferation (Wagner et al., 2014). While we also found a higher proportion of integrations in cancer-related genes as compared to random, this bias was not different from that observed for other highly expressed genes favored by HIV-1. Further, we do not see any differential bias for integration in cancer-related genes in clonally expanded cells compared to single integrations and an overall decrease in the number of integrations in cancer related genes during the course of therapy. Since the number and size of clones increase with time on therapy, the data indicate that integration into cancer genes is unlikely to be a general contributor to the proliferation of infected cells.

Our data show that cART selects for expanded clones and that viremic controllers resemble treated progressors in showing a higher proportion of expanded clones than untreated viremics. cART selects for clonal integrations irrespective of the location in the genome. This is in contrast to single integrations, which are selected against by therapy. cART specifically favors the survival of single integrations in intergenic regions and is biased against genic regions, with an overall half-life for single integrations of 127 months. The half-life of single integrations is not too dissimilar from the current estimate for the latent reservoir, which is believed to decay with a half-life of 44 months on cART (Finzi et al., 1999).

The major outstanding question after the discovery of clonally expanded cells with integrated HIV-1 is whether the virus from these cells contributes to the latent reservoir (Maldarelli et al., 2014; Wagner et al., 2014). Several different independent lines of evidence argue against this idea. First, although the latent reservoir is thought to be contained primarily in resting central memory CD4⁺ T cells (Siliciano and Greene, 2011), we find that clonally expanded viral integrations are found in all three memory T cell compartments. Second, whereas the reservoir appears to decay with time on cART, we find that clonally expanded integrations increase with time and do so irrespective of whether they are found in genes or intergenic regions. In contrast, single integrations in more active parts of the genome, which are more likely to support HIV-1 reactivation, are selected against with time on ART. Finally, all 75 of the clonally expanded proviruses tested were defective, which is in agreement with two examples in the literature (Imamichi et al., 2014; Josefsson et al., 2013). Thus, we conclude that intact virus is not enriched in infected expanded cells. However, we cannot rule out the possibility that a rare clone of cells contains an active virus. Nevertheless, the 90% of all cells bearing integrated proviruses that account for expanded clones of infected cells in cART-treated progressors appear unlikely to be the major source of the rebounding latent reservoir. Instead, the replication-competent reservoir is likely to be contained in the remaining 10% of cells that harbor single integrations that decline with a long half-life on cART (Figure S5).

In conclusion, the data indicate that HIV-1-infected T cells that undergo clonal expansion are able to do so because their proviruses are defective and that the replication-competent reservoir is found in the subset of CD4⁺ T cells that remain quiescent.

EXPERIMENTAL PROCEDURES

CD4⁺ T Cell Isolation for Integration Library Construction

Human samples were collected after signed informed consent in accordance with Institutional Review Board (IRB)-reviewed protocols by all participating institutions. Patients 1, 2, and 3 were selected from the Seattle HIV longitudinal cohort studies at Fred Hutchinson Cancer Research Center. Patients 4, 8, and 9 were recruited from the University of Cologne, and samples were obtained at Rockefeller University (MNU_0628). Patients 5, 6, and 7 were selected from the Rockefeller University HIV-1 antibody therapy clinical trial. Patients 10, 11, 12, and 13 were selected from a group of elite controllers that were followed at the Ragon Institute in Boston.

CD4⁺ T cells were isolated from whole PBMC using anti-CD4 microbeads (Miltenyi Biotec). The percentage of live cells was determined by flow cytometry based on forward and side scatter. Purity of CD4⁺ T cells was determined by labeling isolated cells with anti-human CD3, CD4, CD8, CD19, and HLA-DR and gating on CD3, CD4 double-positive cells. Isolated cells were used for library construction only if purity was >75%. CD4⁺ T cell subsets were isolated by FACS sorting on a BD Aria II by labeling cells with anti-human CD3, CD4, CD8, CD66b, CD335, HLA-DR, CCR7, CD27, and CD45RA. Analysis of CD4⁺ T cell subsets was done by pooling cellular DNA isolated from multiple sorts of the same sample.

Quantitative Viral Outgrowth Assay

Viral outgrowth was performed as previously described. (Laird et al., 2013)

Integration Library

The method for integration library construction was adapted from TC-Seq (Klein et al., 2011).

DNA Preparation

DNA from 0.2–2 million CD4⁺ cells from HIV-1-infected patients was isolated and prepared as previously described (Klein et al., 2011). Fragments were ligated to 200 pmol of annealed linkers (Table S2). Virus sequences were eliminated by digestion with BglII (NEB), and fragments were purified.

Integration Site Amplification

Semi-nested ligation-mediated PCR was performed on DNA. All PCRs were performed using Phusion Polymerase (Thermo). DNA was divided into 700 ng aliquots and subjected to single-primer PCR with biotinylated LTR1 [1 × (98C-1 min) 12 × (98C-15 s, 62C-30 s, 72C-30 s) 1 × (72C-5 min)] (Table S2). Each reaction was spiked with pLinker and subjected to additional cycles of PCR [1 × (98C-1 min) 25 × (98C-15 s, 62C-30 s, 72C-30 s) 1 × (72C-5 min)]. Products of 300–1,000 bp were isolated by agarose gel electrophoresis and magnetic streptavidin bead purification. Semi-nested PCR was performed on the magnetic beads first with a single primer LTR2 (same cycling conditions as above) followed by spiking in pLinker and additional cycles (Table S2). Products of 300–1,000 bp were isolated by gel electrophoresis.

Paired-End Library Preparation

Linkers were digested by *Ascl* such that a six-nucleotide barcode (CGCGCC) was left on the DNA fragments, indicating linker-dependent amplification. Fragments were blunted by End-It DNA Repair Kit (Epicenter), purified with AmPure beads (Agencourt), and ligated to NextFlex paired-end adapters. Adaptor-ligated fragments were enriched by 35 cycles of PCR with NextFlex primers [1 × (98C-1 min) 35 × (98C-15 s, 66C-30 s, 72C-30 s) 1 × (72C-5 min)], and fragments between 300–1,000 bp were isolated by gel electrophoresis. Two or three libraries were mixed in equimolar ratios and sequenced by either 150 bp paired-end sequencing on Illumina MiSeq or 150 bp or 100 bp paired-end sequencing on an Illumina 2500 HiSeq. Data are accessible via NCBI SRA using the accession number: SRP045822.

Computational Analysis

Read Alignment

Paired-end reads were mapped to the HIV-1 sequence (designated as a bait) using BLAT (Kent, 2002) with default settings. Reads that were mapped to the bait without mismatches were checked for the linker barcode in the paired-end read and were mapped to the human genome reference GRCh37/hg19 with Bowtie (Langmead et al., 2009). Only uniquely mapped reads (allowing for up to two mismatches) were used as defined in the best alignment stratum (command line options: -v2 -all -best -strata -m1). Identical reads generated by PCR amplification were merged.

Integration Determination

Once the paired-end reads were properly mapped in the bait and human genome (see above), we determined the integration breakpoint by aligning the remaining nucleotide sequence containing the 3' terminus of the HIV-1 LTR to the human genome using BLAT (default settings). Only uniquely mapped reads up to 1 Kb away from its partner were kept. Adjacent (within 50 nucleotides) putative integration sites were merged. Finally, the 5' end of the paired-end reads were used to deduce the integration and shear position sites in the human genome.

Hotspot Detection

To detect preferred sites of HIV-1 integration genome wide, we subjected our dataset to *hot_scan* software analysis (Silva et al., 2014), which defines hotspots by scan statistics. Hotspots obtained by *hot_scan* were defined using different window widths (100, 200, 500, 1,000, 2,000, 5,000, 10,000, 20,000, 50,000, and 100,000 bp).

Motif Analysis

To determine a consensus motif, 100 bp flanking each integration site was analyzed for the presence of 30 bp consensus sequence using MEME software (Bailey and Elkan, 1994).

Monte Carlo Simulation for Virus Integration and Hotspots

Monte Carlo simulation was conducted by shuffling the genomic locations of all virus integration sites 10,000 times using *bedtools shuffle* utility (Quinlan and Hall, 2010). Then, we compared the observed number with the median number of integrations in the randomized list. We assessed enrichments by p value by counting the frequency of observed events being equal to or higher than the number of randomized events divided by $n = 10,000$.

Statistical Analysis

Proportion test is the standard test for the difference between proportions, also known as a two-proportion z test. We used R's implementation of this via the *prop.test()* function.

Integration Library Verification

To verify our integration sequencing strategy, we constructed two libraries from DNA isolated from un-infected individuals. We recovered 13 sequences that mapped to integration sites. We subtracted these "integration sites" from all libraries before proceeding.

To test the saturation of our method, two separate integration libraries were constructed from identical samples for three patients. We found that both libraries contained the same expanded clonal families, but the majority of single virus integrations were unique to each sample of cells used for library construction. Single viral integrations found in both libraries were less than 1% of observed viral integrations.

PCR Verification

Genomic DNA isolated as described was serially diluted and subjected to nested-PCR using genomic specific primers and primers LTR1 and LTR2 (Table S2) using HotStart Taq Polymerase (QIAGEN) [1 × (98C-14 min) 40 × (98C-30 s, 55C-30 s, 72C-30 s) 1 × (72C-5 min)]. Products were isolated by gel electrophoresis and sequenced directly. Analysis of clones in this manner identified that we underestimate the size of clones by four to five times (data not shown).

CD4⁺ T Cell Subset Sorting

To isolate CD4⁺ subsets, we labeled PBMCs with antibodies to CD45RA, CD4, CD8 CD66b, CCR7, CD335, HLA-DR, CD3, and CD27. We separated T cell subsets by FACS Aria (BD Biosciences) to very high purity (>98%).

Virus Sequencing

5' LTR

5' LTRs from large clones were amplified with nested genomic primers and LTR2Rev (Table S2) using Platinum High Fidelity Taq (Invitrogen) [1 × (98C-14 min) 40 × (98C-30 s, 55C-30 s, 68C-1 min) 1 × (68C-5 min)]. Products were isolated by gel electrophoresis and sequenced directly.

Full-Length Virus

Full-length genomic DNA from infected patients was isolated as described and serially diluted. Each well was filled to a final volume of 50 μ l with PCR reaction mixture (Platinum Taq MasterMix, Invitrogen) and primers to amplify virus from a specific integration site in the genome (Table S2 and Ho et al., 2013) using touchdown cycling to increase specificity. Then, 2 μ l aliquots from the first PCR were subjected to nested genomic PCR and 1% gel electrophoresis. The positive wells were gel purified, and fragments were sequenced directly.

ACCESSION NUMBERS

Data are accessible via NCBI SRA using the accession number SRP045822.

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures, three tables, and one data file and can be found with this article at <http://dx.doi.org/10.1016/j.cell.2015.01.020>.

AUTHOR CONTRIBUTIONS

L.B.C. planned and performed experiments, analyzed the data, and wrote the manuscript. I.T.S. performed all bioinformatics analysis. T.Y.O. performed cancer gene analysis. R.A.R. performed bioinformatics and statistical analysis. E.H.P., G.H.L., and B.H.H. performed SGA and phylogenetic analysis. J.L.C., M.J.M., C.L., F.K., M.C., and B.D.W. selected and provided clinical samples. J.D.S. and R.F.S. performed QVOA and provided extensive discussion. M.J. planned and performed experiments, analyzed the data, and wrote the manuscript. M.C.N. planned experiments, analyzed the data, and wrote the manuscript.

ACKNOWLEDGMENTS

We would like to acknowledge all patients who contributed to this study. We thank Qiao Wang for invaluable discussions; Johannes Scheid for the coordination and preparation of viremic controller samples; Joshua Horwitz for samples during the initial development of integration sequencing; Klara Velinon and Gaelle Breton for FACS sorting; Zoran Jankovic for laboratory support; Rockefeller Genomics Resource Center and New York Genome Center for sequencing; David Chambliss for assistance in obtaining human samples; Arlene Hurley and Gisela Kremer for assistance in patient coordination; and all members of Nussenzweig and Mucida labs for valuable discussion and advice. This work was supported in part by the Bill and Melinda Gates Foundation Collaboration for AIDS Vaccine Discovery Grant OPP 1033115 (to M.C.N.) and CHAVI-ID Award UM1AI100663. This work was also supported in part by grant #8 UL1 TR000043 from the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health (NIH) Clinical and Translational Science Award (CTSA) program. E.H.P., G.H.L., and B.H.H. are supported by UM1 AI100645 and R37 AI 066998. M.C.N., R.F.S., and B.D.W. are Howard Hughes Medical Institute Investigators. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Received: August 27, 2014
Revised: December 18, 2014
Accepted: January 12, 2015
Published: January 29, 2015

REFERENCES

Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings / International*

Conference on Intelligent Systems for Molecular Biology; ISMB International Conference on Intelligent Systems for Molecular Biology 2, 28–36.

Berry, C.C., Gillet, N.A., Melamed, A., Gormley, N., Bangham, C.R., and Bushman, F.D. (2012). Estimating abundances of retroviral insertion sites from DNA fragment length data. *Bioinformatics* 28, 755–762.

Brady, T., Agosto, L.M., Malani, N., Berry, C.C., O'Doherty, U., and Bushman, F. (2009). HIV integration site distributions in resting and activated CD4+ T cells infected in culture. *AIDS* 23, 1461–1471.

Buzon, M.J., Sun, H., Li, C., Shaw, A., Seiss, K., Ouyang, Z., Martin-Gayo, E., Leng, J., Henrich, T.J., Li, J.Z., et al. (2014). HIV-1 persistence in CD4+ T cells with stem cell-like properties. *Nat. Med.* 20, 139–142.

Chomont, N., El-Far, M., Ancuta, P., Trautmann, L., Procopio, F.A., Yassine-Diab, B., Boucher, G., Boulassel, M.R., Ghattas, G., Brechley, J.M., et al. (2009). HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nat. Med.* 15, 893–900.

Chun, T.W., Carruth, L., Finzi, D., Shen, X., DiGiuseppe, J.A., Taylor, H., Hermankova, M., Chadwick, K., Margolick, J., Quinn, T.C., et al. (1997). Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature* 387, 183–188.

Chun, T.W., Engel, D., Berrey, M.M., Shea, T., Corey, L., and Fauci, A.S. (1998). Early establishment of a pool of latently infected, resting CD4(+) T cells during primary HIV-1 infection. *Proc. Natl. Acad. Sci. USA* 95, 8869–8873.

Craigie, R., and Bushman, F.D. (2012). HIV DNA integration. *Cold Spring Harbor perspectives in medicine* 2, a006890.

Finzi, D., Hermankova, M., Pierson, T., Carruth, L.M., Buck, C., Chaisson, R.E., Quinn, T.C., Chadwick, K., Margolick, J., Brookmeyer, R., et al. (1997). Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science* 278, 1295–1300.

Finzi, D., Blankson, J., Siliciano, J.D., Margolick, J.B., Chadwick, K., Pierson, T., Smith, K., Lisiewicz, J., Lori, F., Flexner, C., et al. (1999). Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nat. Med.* 5, 512–517.

Han, Y., Lassen, K., Monie, D., Sedaghat, A.R., Shimoji, S., Liu, X., Pierson, T.C., Margolick, J.B., Siliciano, R.F., and Siliciano, J.D. (2004). Resting CD4+ T cells from human immunodeficiency virus type 1 (HIV-1)-infected individuals carry integrated HIV-1 genomes within actively transcribed host genes. *J. Virol.* 78, 6122–6133.

Ho, Y.C., Shan, L., Hosmane, N.N., Wang, J., Laskey, S.B., Rosenbloom, D.I., Lai, J., Blankson, J.N., Siliciano, J.D., and Siliciano, R.F. (2013). Replication-competent noninduced proviruses in the latent reservoir increase barrier to HIV-1 cure. *Cell* 155, 540–551.

Holman, A.G., and Coffin, J.M. (2005). Symmetrical base preferences surrounding HIV-1, avian sarcoma/leukosis virus, and murine leukemia virus integration sites. *Proc. Natl. Acad. Sci. USA* 102, 6103–6107.

Ikeda, T., Shibata, J., Yoshimura, K., Koito, A., and Matsushita, S. (2007). Recurrent HIV-1 integration at the BACH2 locus in resting CD4+ T cell populations during effective highly active antiretroviral therapy. *J. Infect. Dis.* 195, 716–725.

Imamichi, H., Natarajan, V., Adelsberger, J.W., Rehm, C.A., Lempicki, R.A., Das, B., Hazen, A., Imamichi, T., and Lane, H.C. (2014). Lifespan of effector memory CD4+ T cells determined by replication-incompetent integrated HIV-1 provirus. *AIDS* 28, 1091–1099.

Janovitz, T., Klein, I.A., Oliveira, T., Mukherjee, P., Nussenzweig, M.C., Sadelain, M., and Falck-Pedersen, E. (2013). High-throughput sequencing reveals principles of adeno-associated virus serotype 2 integration. *J. Virol.* 87, 8559–8568.

Jordan, A., Defechereux, P., and Verdin, E. (2001). The site of HIV-1 integration in the human genome determines basal transcriptional activity and response to Tat transactivation. *EMBO J.* 20, 1726–1738.

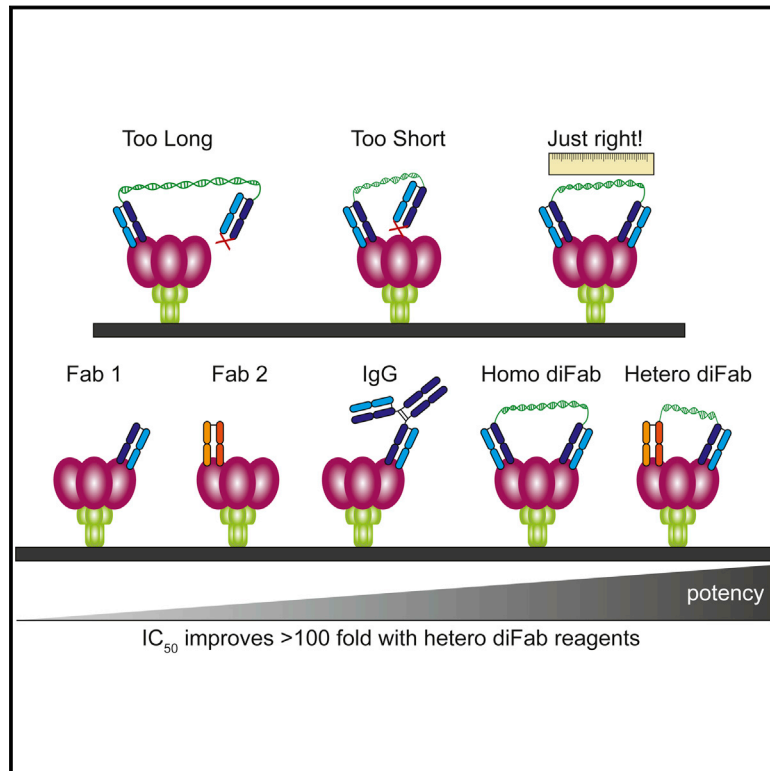
Jordan, A., Bisgrove, D., and Verdin, E. (2003). HIV reproducibly establishes a latent infection after acute infection of T cells in vitro. *EMBO J.* 22, 1868–1877.

Josefsson, L., von Stockenström, S., Faria, N.R., Sinclair, E., Bacchetti, P., Killian, M., Epling, L., Tan, A., Ho, T., Lemey, P., et al. (2013). The HIV-1 reservoir in

- eight patients on long-term suppressive antiretroviral therapy is stable with few genetic changes over time. *Proc. Natl. Acad. Sci. USA* **110**, E4987–E4996.
- Kent, W.J. (2002). BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664.
- Klein, I.A., Resch, W., Jankovic, M., Oliveira, T., Yamane, A., Nakahashi, H., Di Virgilio, M., Bothmer, A., Nussenzweig, A., Robbiani, D.F., et al. (2011). Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. *Cell* **147**, 95–106.
- Laird, G.M., Eisele, E.E., Rabi, S.A., Lai, J., Chioma, S., Blankson, J.N., Siliciano, J.D., and Siliciano, R.F. (2013). Rapid quantification of the latent reservoir for HIV-1 using a viral outgrowth assay. *PLoS Pathog.* **9**, e1003398.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25.
- Lewinski, M.K., Yamashita, M., Emerman, M., Ciuffi, A., Marshall, H., Crawford, G., Collins, F., Shinn, P., Leipzig, J., Hannehalli, S., et al. (2006). Retroviral DNA integration: viral and cellular determinants of target-site selection. *PLoS Pathog.* **2**, e60.
- Maldarelli, F., Wu, X., Su, L., Simonetti, F.R., Shao, W., Hill, S., Spindler, J., Ferris, A.L., Mellors, J.W., Kearney, M.F., et al. (2014). HIV latency. Specific HIV integration sites are linked to clonal expansion and persistence of infected cells. *Science* **345**, 179–183.
- Mitchell, R.S., Beitzel, B.F., Schroder, A.R., Shinn, P., Chen, H., Berry, C.C., Ecker, J.R., and Bushman, F.D. (2004). Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol.* **2**, E234.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842.
- Rose, P.P., and Korber, B.T. (2000). Detecting hypermutations in viral sequences with an emphasis on G → A hypermutation. *Bioinformatics* **16**, 400–401.
- Schröder, A.R., Shinn, P., Chen, H., Berry, C., Ecker, J.R., and Bushman, F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* **110**, 521–529.
- Serrao, E., Krishnan, L., Shun, M.C., Li, X., Cherepanov, P., Engelman, A., and Maertens, G.N. (2014). Integrase residues that determine nucleotide preferences at sites of HIV-1 integration: implications for the mechanism of target DNA binding. *Nucleic Acids Res.* **42**, 5164–5176.
- Sherrill-Mix, S., Lewinski, M.K., Famiglietti, M., Bosque, A., Malani, N., Ocwieja, K.E., Berry, C.C., Looney, D., Shan, L., Agosto, L.M., et al. (2013). HIV latency and integration site placement in five cell-based models. *Retrovirology* **10**, 90.
- Siliciano, R.F., and Greene, W.C. (2011). HIV latency. *Cold Spring Harbor perspectives in medicine* **1**, a007096.
- Silva, I.T., Rosales, R.A., Holanda, A.J., Nussenzweig, M.C., and Jankovic, M. (2014). Identification of chromosomal translocation hotspots via scan statistics. *Bioinformatics* **30**, 2551–2558.
- Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., Jr., and Kinzler, K.W. (2013). Cancer genome landscapes. *Science* **339**, 1546–1558.
- Wagner, T.A., McKernan, J.L., Tobin, N.H., Tapia, K.A., Mullins, J.I., and Frenkel, L.M. (2013). An increasing proportion of monotypic HIV-1 DNA sequences during antiretroviral treatment suggests proliferation of HIV-infected cells. *J. Virol.* **87**, 1770–1778.
- Wagner, T.A., McLaughlin, S., Garg, K., Cheung, C.Y., Larsen, B.B., Styrchak, S., Huang, H.C., Edlefsen, P.T., Mullins, J.I., and Frenkel, L.M. (2014). HIV latency. Proliferation of cells with HIV integrated into cancer genes contributes to persistent infection. *Science* **345**, 570–573.
- Wang, G.P., Ciuffi, A., Leipzig, J., Berry, C.C., and Bushman, F.D. (2007). HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome Res.* **17**, 1186–1194.
- Wong, J.K., Hezareh, M., Günthard, H.F., Havlir, D.V., Ignacio, C.C., Spina, C.A., and Richman, D.D. (1997). Recovery of replication-competent HIV despite prolonged suppression of plasma viremia. *Science* **278**, 1291–1295.
- Wu, X., Li, Y., Crise, B., Burgess, S.M., and Munroe, D.J. (2005). Weak palindromic consensus sequences are a common feature found at the integration target sites of many retroviruses. *J. Virol.* **79**, 5211–5214.
- Zhao, M., Sun, J., and Zhao, Z. (2013). TSGene: a web resource for tumor suppressor genes. *Nucleic Acids Res.* **41**, D970–D976.

Intra-Spike Crosslinking Overcomes Antibody Evasion by HIV-1

Graphical Abstract



Authors

Rachel P. Galimidi, Joshua S. Klein, ..., Anthony P. West Jr., Pamela J. Bjorkman

Correspondence

bjorkman@caltech.edu

In Brief

While antibodies generally neutralize viruses by bivalent binding to neighboring virion spikes, HIV-1 virion and spike architectures generally prohibit both inter- and intra-spike crosslinking by naturally occurring antibodies. Newly engineered antibody-based molecules, rationally designed for high-avidity intraspine binding, overcome this barrier and exhibit >100-fold increases in HIV-1 neutralization potencies.

Highlights

- Low density of HIV spikes interferes with bivalent binding by antibodies
- Monovalent binding contributes to antibody vulnerability to HIV-1 mutation
- ≥ 100 -fold increases in potency achieved by intra-spike crosslinking
- Ideal anti-HIV-1 antibody therapeutics bind to single spikes bivalently



Intra-Spike Crosslinking Overcomes Antibody Evasion by HIV-1

Rachel P. Galimidi,¹ Joshua S. Klein,^{1,6} Maria S. Politzer,¹ Shiyu Bai,^{1,5} Michael S. Seaman,² Michel C. Nussenzweig,^{3,4} Anthony P. West, Jr.,¹ and Pamela J. Bjorkman^{1,4,*}

¹Division of Biology and Biological Engineering, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA

²Beth Israel Deaconess Medical Center, Boston, MA 02215, USA

³Laboratory of Molecular Immunology, The Rockefeller University, New York, NY 10065, USA

⁴Howard Hughes Medical Institute

⁵Present address: Case Western Reserve University School of Medicine, 2109 Adelbert Road, Cleveland, OH 44106, USA

⁶Present address: Google Inc., 1600 Amphitheatre Parkway, Mountain View, CA 94043, USA

*Correspondence: bjorkman@caltech.edu

<http://dx.doi.org/10.1016/j.cell.2015.01.016>

SUMMARY

Antibodies developed during HIV-1 infection lose efficacy as the viral spike mutates. We postulated that anti-HIV-1 antibodies primarily bind monovalently because HIV's low spike density impedes bivalent binding through inter-spike crosslinking, and the spike structure prohibits bivalent binding through intra-spike crosslinking. Monovalent binding reduces avidity and potency, thus expanding the range of mutations permitting antibody evasion. To test this idea, we engineered antibody-based molecules capable of bivalent binding through intra-spike crosslinking. We used DNA as a "molecular ruler" to measure intra-epitope distances on virion-bound spikes and construct intra-spike crosslinking molecules. Optimal bivalent reagents exhibited up to 2.5 orders of magnitude increased potency (>100-fold average increases across virus panels) and identified conformational states of virion-bound spikes. The demonstration that intra-spike crosslinking lowers the concentration of antibodies required for neutralization supports the hypothesis that low spike densities facilitate antibody evasion and the use of molecules capable of intra-spike crosslinking for therapy or passive protection.

INTRODUCTION

The HIV-1 envelope (Env) spike trimer, a trimer of gp120 and gp41 subunits, is the only target of neutralizing antibodies. The spike utilizes antibody-evasion strategies, including mutation, glycan shielding, and conformational masking (West et al., 2014). Although important, these features are not unique to HIV-1; other viruses employing these strategies elicit IgG antibody responses that provide sterilizing immunity or viral clearance. A potentially unique antibody-evasion strategy for HIV-1 involves hindering IgGs from using both antigen-binding fragments (Fabs) to bind bivalently to spikes (Klein and Bjorkman,

2010; Mouquet et al., 2010). This is accomplished by the small number and low density of Env spikes (Chertova et al., 2002; Liu et al., 2008; Zhu et al., 2006), which prevent most IgGs from inter-spike crosslinking (bivalent binding between spikes), and the architecture of the Env trimer, which impedes intra-spike crosslinking (bivalent binding within a spike trimer) (Klein et al., 2009; Luftig et al., 2006).

On a typical virus with closely spaced envelope spikes, an IgG antibody can bind using both Fabs to crosslink neighboring spikes, leading to a nearly irreversible antibody-antigen interaction (Mattes, 2005). Avidity effects from bivalent binding of IgG antibodies have been shown to be critical for neutralization of many viruses, including polio and influenza (Icenogle et al., 1983; Schofield et al., 1997). By contrast, the small number of spikes (~14) present on the surface of HIV-1 (Chertova et al., 2002; Liu et al., 2008; Zhu et al., 2006) impedes simultaneous engagement of both antibody combining sites (Klein and Bjorkman, 2010; Mouquet et al., 2010)—most spikes are separated by distances that far exceed the ~15 nm reach of the two Fab arms of an IgG (Liu et al., 2008; Zhu et al., 2006) (Figure 1A). Inter-spike crosslinking might still be possible if spikes could freely diffuse within the viral membrane. However, cryo-electron tomography of HIV-1 (Zhu et al., 2006) and evidence for interactions between the cytoplasmic tail of gp41 and the matrix protein of HIV (Bhatia et al., 2009; Crooks et al., 2008; Yu et al., 1992) suggest that a virion's spike distribution is likely to be relatively static over timescales relevant to neutralization. Taken together, the mechanisms to hinder inter- and intra-spike crosslinking imply that most anti-HIV-1 IgGs bind monovalently to virions.

It seems an unlikely coincidence that HIV-1, among the most adept of viruses at evading antibody-mediated neutralization, has an unusually low density of surface envelope spikes with restricted mobility, as well as an unusually high mutation rate. We speculated that HIV-1 evolved a low spike density to hinder bivalent binding by antibodies (Klein and Bjorkman, 2010) and postulated that the combination of predominantly monovalent IgG binding and HIV-1's rapid mutation rate creates an additional effective antibody evasion strategy (Klein and Bjorkman, 2010). If the affinity between an IgG Fab and a viral spike is high enough, monovalent IgG binding to a virion should not, in and of itself, hinder or prevent viral neutralization. Thus, affinity-matured anti-Env IgGs raised against a particular strain of virus can

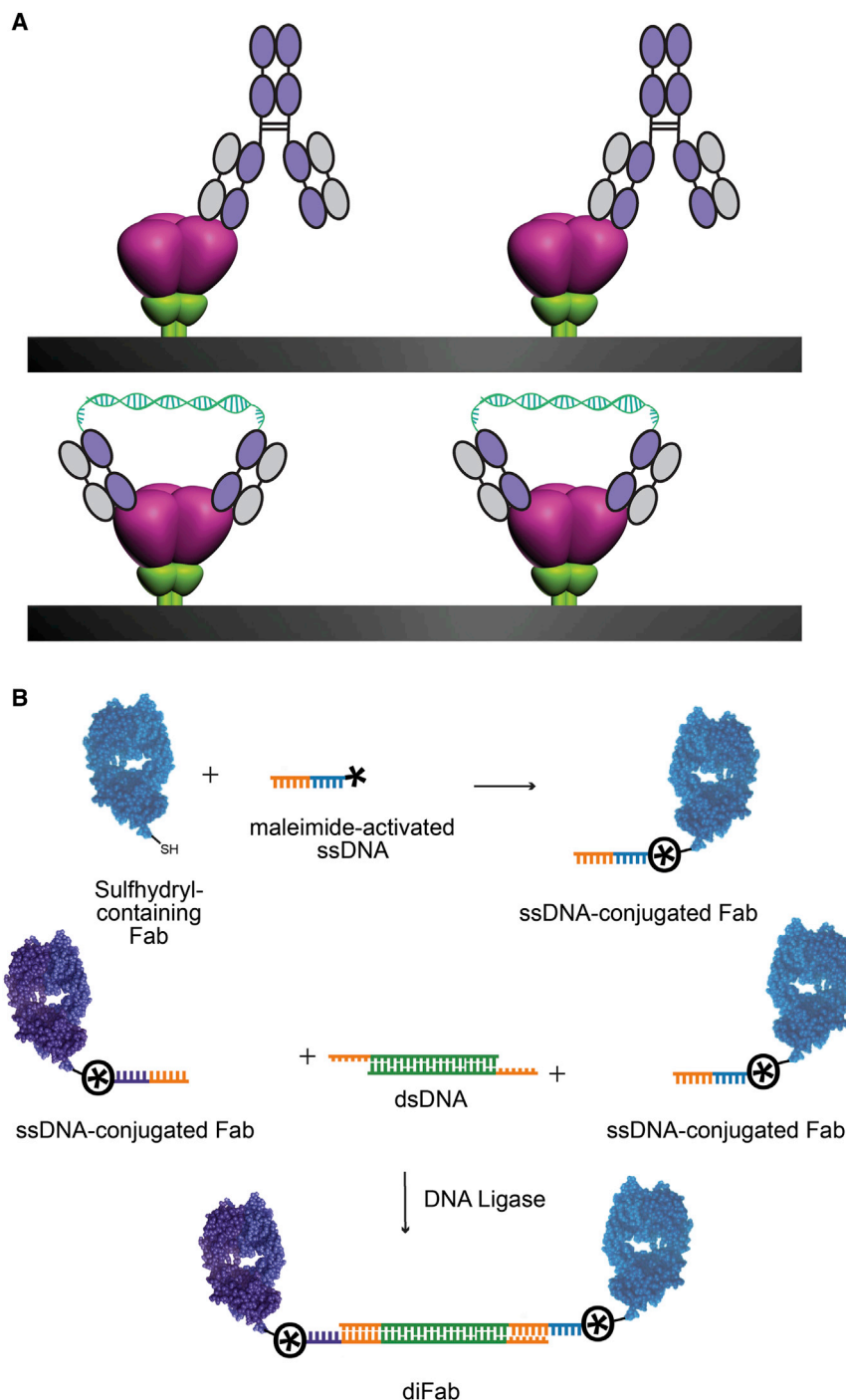


Figure 1. IgG and diFab Reagents Binding to Viral Spikes

(A) Top: IgG binding monovalently to spikes on HIV-1 surfaces, which include a small number (~14) and low density of Env (Chertova et al., 2002; Liu et al., 2008; Zhu et al., 2006). Bottom: homo-diFab reagent binding bivalently to HIV-1 Env by intra-spike crosslinking. Schematic representations of Env adapted from figures in Liu et al. (2008).

(B) Schematic of method used to produce homo- and hetero-diFabs.

See also Figure S1.

a high density of Env spikes (Liljeroos et al., 2013). Palivizumab Fabs with fast off-rates/low affinities exhibited 2–3 log improvements in neutralization potencies when converted to bivalent IgGs, and the potencies of the IgGs were not affected by mutations that increased the off-rates of their corresponding monovalent Fabs by >100-fold (Wu et al., 2005), illustrating the importance of avidity for IgGs with weak or moderate affinity Fabs. However, high affinity/slow off-rate palivizumab Fabs were equally as potent as their IgG counterparts, which could bind bivalently to RSV through inter-spike crosslinking. In the palivizumab example, binding and neutralization potencies were evaluated for a single strain of virus and antibodies. In the case of HIV-1, we are interested in the effects of mutations in the virus on binding of the same antibody, but the effects of mutation are expected to be similar. Thus, we postulate that avidity effects through bivalent binding can serve as a buffer to dampen the effects of viral mutations on neutralization potencies of IgGs.

This line of reasoning suggests that bivalent HIV-1 binders would be optimal for passive prevention or immunotherapy, but because inter-spike distances are not constant, even on a single virion, it is not possible to engineer reagents that could consistently accomplish inter-spike crosslinking. In contrast, reagents that can bind bivalently to a single trimeric spike would function independently of both spike density and distribution (Pace et al., 2013).

To test the idea that intra-spike crosslinking results in increased neutralization potency, we used molecular rulers to map epitopes on virion-bound HIV-1 spikes and created molecules designed to synergize through bivalent interactions within single Env trimers (Figure 1A). We developed methods to produce multiple combinations of Env binders separated by different distances by

effectively neutralize autologous virus (Klein et al., 2013; West et al., 2014). However, upon mutation of an antibody epitope on Env, the low affinity of the monovalent Fab-antigen interaction would result in either complete loss of neutralization or neutralization only at very high concentrations. These concepts are illustrated by comparisons of binding and neutralization for variants of IgG and Fab forms of palivizumab, a neutralizing IgG against respiratory syncytial virus (RSV) (Wu et al., 2005), a virus with

attaching broadly neutralizing antibody (bNAb) Fabs to variable-length double-stranded DNA (dsDNA) (Figures 1B and S1). We chose dsDNA as a linker because its long persistence length (460–500 Å [Bednar et al., 1995] compared with ~30 Å for peptides [Zhou, 2004]) permits its use as a molecular ruler with 3.4 Å/base pair (bp) increments. Here, we show that homo- and hetero-diFabs joined by optimal-length dsDNA bridges can achieve neutralization potency increases of two to three orders of magnitude and provide evidence that the synergy results from intra-spike crosslinking. Upon determining the optimal distances between Env trimer-bound Fabs, we show that it is possible to convert the dsDNA bridge to a protein linker to create a protein-based reagent with similar synergistic properties. These results illustrate the importance of avidity in antibody-pathogen interactions, elucidate mechanisms by which HIV-1 evades the host immune system, and are relevant to the choice of potential protein therapeutics to be delivered to prevent or treat HIV-1 infections.

RESULTS

Homo-diFabs Exhibit Length-Dependent Avidity Effects Consistent with Intra-Spike Crosslinking

Fabs were modified to contain a free thiol and then conjugated to maleimide-activated single-stranded DNA (ssDNA) (Figure 1B). Different lengths of dsDNA (designed to lack secondary structures [Zadeh et al., 2011]) (Extended Experimental Procedures and Table S6) were annealed with and ligated to the ssDNA-Fab conjugates to create homo- or hetero-diFabs, in which the two Fabs were the same or different, respectively. Dynamic light scattering confirmed that conjugates with longer DNA bridges were more extended (Figure 2A), supporting the use of dsDNA as a ruler. Inter-Fab distances calculated from dsDNA lengths were regarded as approximate because the DNA linkers included short regions of ssDNA (persistence length 22 Å) [Chi et al., 2013] to permit orientational flexibility.

We first determined the optimal dsDNA linker for a homo-diFab constructed from 3BNC60, a bNAb against the CD4 binding site (CD4bs) on the gp120 subunit of Env [Scheid et al., 2011], by evaluating homo-diFabs with different dsDNA lengths using *in vitro* neutralization assays. The 50% inhibitory concentrations (IC₅₀s) against HIV-1 strain 6535.3 depended on the dsDNA length, with the most potent homo-diFab containing a bridge of 62 bp (211 Å) (Figures 2B and S2). This length is close to the predicted distance (~198 Å) between the C termini of adjacent 3BNC60 Fabs bound to the open structure of an HIV-1 trimer [Merk and Subramaniam, 2013] (Figures 3 and S3). Bridge lengths of ~60 bp also exhibited the best potencies for 3BNC60 homo-diFabs against DU172.17 HIV-1 and for homo-diFabs constructed from VRC01 [Wu et al., 2010], a related CD4bs bNAb (Figure S2). The ~100-fold increased potency of 3BNC60-62bp-3BNC60 compared with 3BNC60 IgG against HIV-1 6535.3 (Figure 2B) suggested synergy resulting from avidity effects due to bivalent binding. The bivalent interaction likely resulted from intra-spike crosslinking rather than inter-spike crosslinking because the latter should not manifest with a sharp length dependence because inter-spike distances are variable within and between virions [Liu et al., 2008; Zhu et al., 2006].

To formally assess the extent to which inter-spike crosslinking could contribute to synergy, we evaluated homo-diFabs constructed from the V1V2 loop-specific bNAb PG16 [Walker et al., 2009], which cannot crosslink within a single spike because only one anti-V1V2 Fab binds per Env trimer [Julien et al., 2013b]. PG16 homo-diFabs with different dsDNA bridges did not exhibit length-dependent neutralization profiles against strain 6535.3 (Figure 2B) and other viral strains (Figure S2D). However, increased potencies were observed for PG16 homo-diFabs with ≥ 70 bp or 80 bp (≥ 248 Å or 272 Å) bridges, perhaps reflecting increased inter-spike crosslinking with longer separation distances (Figures 2B and S2D).

Comparison of Homo-diFabs that Can or Cannot Exhibit Intra-Spike Crosslinking

To evaluate the potential for intra-spike crosslinking across different viral strains, we compared homo-diFabs designed to be capable (b12 and 3BNC60) or incapable (PG16) of intra-spike crosslinking (Figure 2C). To minimize inter-spike crosslinking, the homo-diFabs were constructed with 60–62 bp bridges. The b12-60bp-b12 homo-diFab exhibited increased potency compared with b12 IgG in 21 of 25 strains in a cross-clade panel of primary HIV-1, with potency increases ≥ 10-fold for 16 strains and a geometric mean potency increase of 22-fold. 3BNC60-62bp-3BNC60 showed even more consistent synergy, being more potent than 3BNC60 IgG against all 25 strains tested, with ≥ 10-fold increases for 20 strains and a mean increase of 19-fold. By contrast, the PG16-60bp-PG16 homo-diFab showed potency increases compared with PG16 IgG against only six strains, with relatively small (2- to 7-fold) increases in five strains and an overall 2.8-fold mean potency change.

Hetero-diFabs Exhibit Dramatic Potency Increases Consistent with Intra-Spike Crosslinking

To determine whether heterotypic bivalent binding can produce synergy and to measure distances between epitopes, we used dsDNA to link Fabs recognizing different epitopes on gp120. We first evaluated hetero-diFabs constructed with Fabs from V1V2 (PG16 or PG9) [Walker et al., 2009] and CD4bs (b12 or 3BNC60) [Roben et al., 1994; Scheid et al., 2011] bNAbs linked with 60 bp dsDNA bridges. PG16-60bp-b12 hetero-diFabs were evaluated in neutralization assays against HIV-1 strains SC4226618 (more sensitive to b12 than PG16) and CAP210 (more sensitive to PG16 than b12). According to the model being tested, in the absence of synergistic binding; i.e., when only one Fab can bind to a spike at a time, a hetero-diFab would be no more potent than a non-covalent mixture of the dsDNA and the two Fabs against each viral strain, whereas synergistic binding would result in avidity effects exhibited by increased potency of the hetero-diFab. For both viral strains, the PG16-60bp-b12 hetero-diFab was ~10-fold more potent than the mixture of Fabs plus dsDNA or the more potent of the two Fabs alone (Figures 4 and S4). To more systematically explore potential synergy, we evaluated PG16-60bp-b12 against a 25 member panel of HIV-1 strains, finding synergistic effects (between 2- and 145-fold more potent than the corresponding non-covalent mixture for most strains; geometric mean improvement of 4.7-fold) (Table S1). When Fabs from PG16 or PG9 were combined with

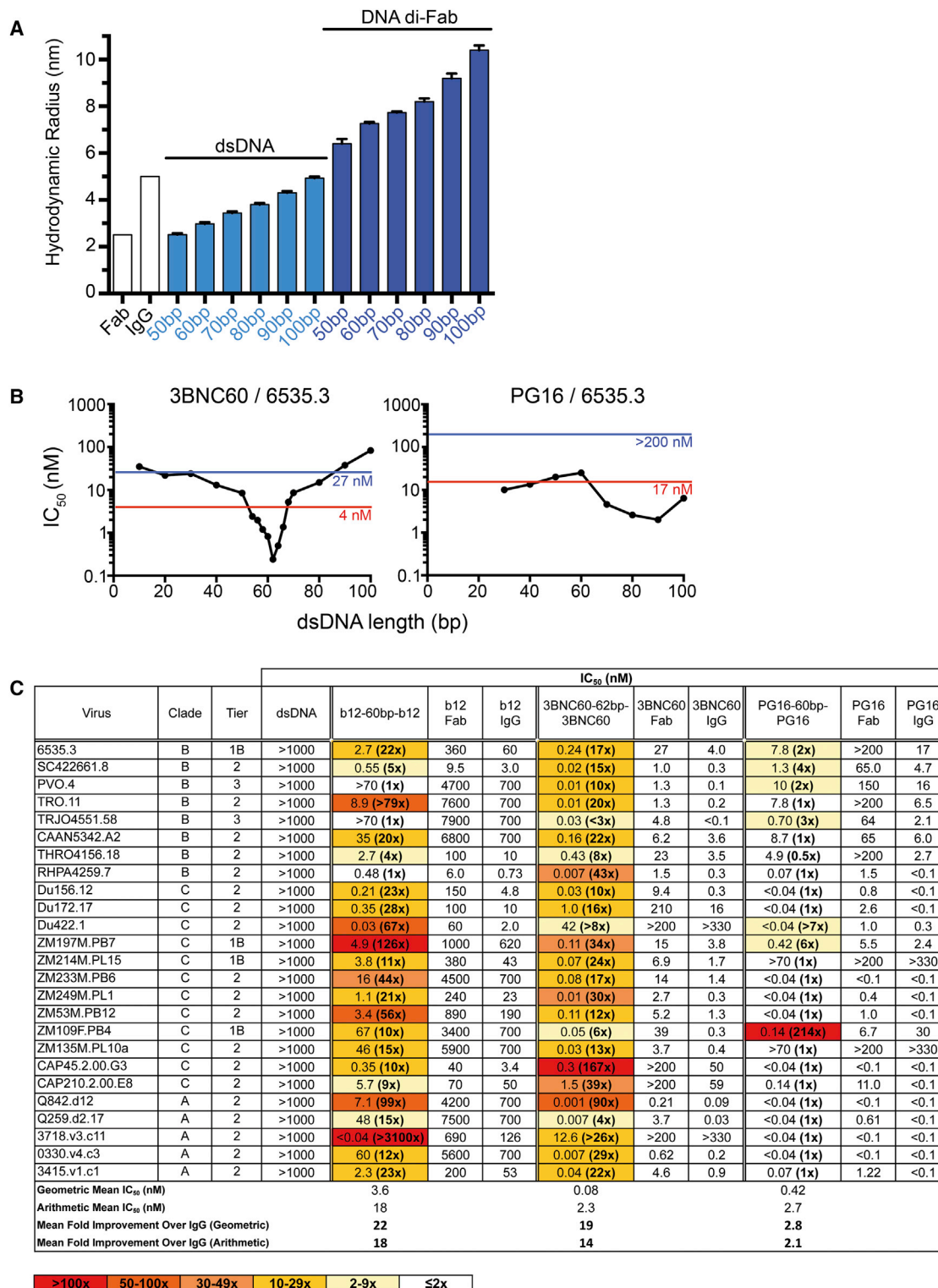


Figure 2. Characterization of Homo-diFabs

(A) Dynamic light-scattering measurements of hydrodynamic radii for IgG and Fab proteins, different lengths of dsDNA alone, and di-Fabs with different dsDNA linkers.

(B) Effects of dsDNA bridge length on neutralization potencies of 3BNC60 and PG16 homo-diFabs against the Tier 1B HIV-1 strain 6535.3. Neutralization IC₅₀s are plotted against the length of the dsDNA linker. IC₅₀s for the parent IgG and Fab are indicated as red and blue lines, respectively.

(legend continued on next page)

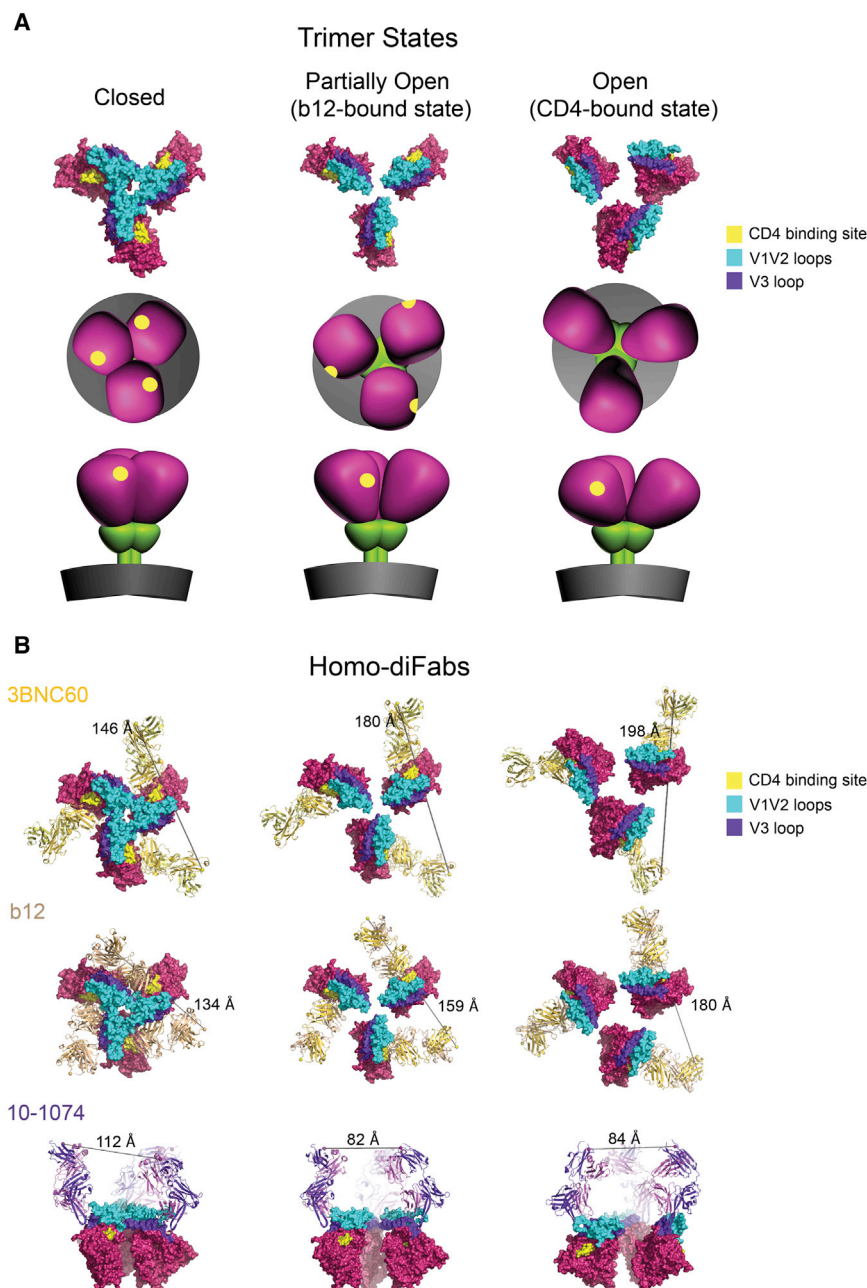


Figure 3. Comparison of Intra-Spike Distances for Three Conformations Found for Virion-Associated HIV-1 Env Spike Trimers

(A) Three conformations of Env trimers shown as surface representations (top row: gp120 coordinates only) and schematically (bottom two rows). Schematic representations of Env trimers adapted from figures in Liu et al. (2008). Env spikes are shown as seen from above (top and middle rows) and the side (bottom row). V1V2 loops are cyan, V3 loops are purple, the CD4 binding site is yellow, the remainder of gp120 is maroon, gp41 is green, and the membrane bilayer is gray. The closed structure (PDB code 4NCO) was observed for unliganded trimers (Liu et al., 2008) and trimers associated with Fabs from potent VRC01-like (PVL) antibodies (Lyumkis et al., 2013; Merk and Subramaniam, 2013). The open structure was observed for trimers associated with CD4 or the Fab from the CD4-induced antibody 17b (Merk and Subramaniam, 2013; Tran et al., 2012) (coordinates obtained from S. Subramaniam). The partially open structure was observed for trimers associated with the Fab from b12 (Liu et al., 2008; Merk and Subramaniam, 2013) (PDB code 3DNL). (B) Measured distances between homo-diFabs bound to HIV-1 trimer structures. Fabs from the indicated bNAbs are shown bound to the gp120 portions of Env in the three conformation shown in (A). Fabs are shown as ribbons; gp120 subunits are shown as surface representations with V1V2 loops in cyan, V3 in purple, the CD4 binding site in yellow, and the remainder of gp120 in maroon. The distance between the Cys233_{heavy chain} carbon- α atoms of adjacent bound Fabs is indicated by a gray line as an approximation of an optimal length for a dsDNA bridge attached to Cys233_{heavy chain}. Assuming 3-fold symmetry of trimers, only one distance is possible for bound 3BNC60, b12, and 10-1074 homo-diFabs. See also Figure S3.

a more potent CD4bs-recognizing bNAb (3BNC60), the resulting hetero-diFabs exhibited greater synergy—several examples of >150-fold improvement for PG16-60bp-3BNC60 and PG9-60bp-3BNC60 and geometric mean potency improvements of 29- and 68-fold, respectively (Figure 4 and Tables S2 and S3). Other hetero-diFabs, constructed with combinations of Fabs recognizing the CD4bs (3BNC60 [Scheid et al., 2011]), the

gp120 V3 loop (10-1074 [Mouquet et al., 2012]), and a gp41 epitope (10E8 [Huang et al., 2012]), also showed synergistic effects (Figure 4 and Table S4), and a 3BNC60-60bp-b12 hetero-diFab exhibited up to 660-fold synergy and a geometric mean potency increase of 90-fold (Figure 4 and Table S5). In contrast, analogous IgG heterodimers, constructed with two different Fabs linked to a single Fc (Schaefer et al., 2011), did not show synergy when evaluated against the same viruses, demonstrating that synergistic effects required optimal separation distances that permitted each Fab to achieve its specific binding orientation (Figure S4 and Tables S1, S2, S3, S4, and S5). We conclude that hetero-diFabs can

(C) Neutralization of primary HIV-1 strains by b12 and PG16 homo-diFabs, each constructed with a 60 bp dsDNA bridge. IC₅₀s are reported for the homo-diFabs, the parental Fabs and IgGs, and dsDNA alone. As a measure of potential synergy, the molar ratio of the IC₅₀ values for the IgG and the homo-diFab is listed for each strain in parentheses beside the IC₅₀ for the homo-diFab.

See also Figure S2.

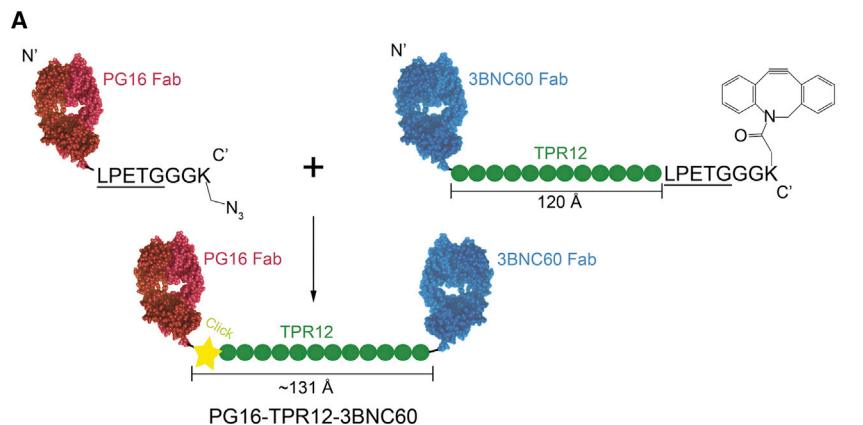
Virus	Clade	Tier	IC ₅₀ (nM)				
			PG16-40bp-3BNC60	PG16-50bp-3BNC60	PG16-60bp-3BNC60	PG9-60bp-3BNC60	3BNC60-60bp-b12
6535.3	B	1B	0.31 (351x)	0.23 (474x)	0.66 (165x)	<0.03 (>767x)	0.22 (>172x)
SC422661.8	B	2	0.021 (86x)	0.01 (180x)	0.05 (36x)	<0.03 (>47x)	<0.04 (>30x)
PVO.4	B	3	0.04 (40x)	0.03 (53x)	0.04 (40x)	<0.03 (>40x)	0.04 (50x)
TRO.11	B	2	0.03 (67x)	0.03 (67x)	0.05 (40x)	<0.03 (>40x)	<0.04 (>55x)
TRJO4551.58	B	3	0.04 (145x)	0.03 (193x)	0.1 (58x)	<0.03 (>187x)	<0.04 (>165x)
CAAN5342.A2	B	2	0.22 (118x)	0.17 (153x)	0.5 (52x)	200 (0.08x)	0.3 (72x)
THRO4156.18	B	2	0.59 (80x)	0.38 (124x)	1.2 (39x)	22 (1x)	0.52 (23x)
RHPA4259.7	B	2	0.007 (200x)	0.007 (200x)	0.02 (70x)	<0.03 (>20x)	<0.04 (>70x)
Du156.12	C	2	0.007 (114x)	0.007 (114x)	0.04 (20x)	<0.03 (>53x)	<0.04 (>205x)
Du172.17	C	2	0.06 (70x)	0.06 (70x)	0.36 (12x)	<0.03 (>273x)	0.22 (117x)
Du422.1	C	2	0.36 (14x)	0.41 (12x)	0.25 (20x)	<0.03 (>160x)	0.15 (13x)
ZM197M.PB7	C	1B	0.06 (107x)	0.06 (107x)	0.18 (35x)	19 (0.5x)	0.22 (91x)
ZM214M.PL15	C	1B	0.001 (9000x)	0.002 (4500x)	0.007 (1280x)	78 (0.1x)	0.15 (81x)
ZM233M.PB6	C	2	0.001 (100x)	0.001 (100x)	0.007 (14x)	<0.03 (>3x)	0.07 (320x)
ZM249M.PL1	C	2	0.01 (60x)	0.01 (60x)	0.03 (20x)	<0.03 (>27x)	<0.04 (>155x)
ZM53M.PB12	C	2	0.04 (25x)	0.03 (33x)	0.13 (8x)	<0.03 (>33x)	0.07 (149x)
ZM109F.PB4	C	1B	0.03 (133x)	0.03 (133x)	0.12 (33x)	<0.03 (>280x)	0.07 (660x)
ZM135M.PL10a	C	2	0.06 (83x)	0.06 (83x)	0.14 (36x)	<0.03 (>140x)	0.07 (69x)
CAP45.2.00.G3	C	2	<0.0007 (>143x)	<0.0007 (>143x)	0.003 (33x)	<0.03 (>3x)	0.07 (29x)
CAP210.2.00.E8	C	2	0.05 (52x)	0.02 (130x)	0.11 (24x)	<0.03 (>140x)	1.3 (>154x)
Q842.d12	A	2	0.001 (200x)	0.001 (200x)	0.007 (29x)	<0.03 (>7x)	<0.04 (>15x)
Q259.d2.17	A	2	0.01 (80x)	0.01 (80x)	0.03 (27x)	<0.03 (>33x)	<0.04 (>110x)
3718.v3.c11	A	2	0.03 (13x)	0.03 (13x)	0.12 (3x)	<0.03 (>27x)	5.67 (35x)
0330.v4.c3	A	2	0.002 (50x)	0.001 (100x)	0.007 (14x)	<0.03 (>7x)	<0.04 (>40x)
3415.v1.c1	A	2	0.01 (120x)	0.01 (120x)	0.05 (24x)	<0.03 (>67x)	0.07 (94x)
Geometric Mean IC ₅₀ (nM)			0.022	0.020	0.075	0.038	0.091
Arithmetic Mean IC ₅₀ (nM)			0.083	0.068	0.24	0.050	0.36
Mean Fold Improvement Over Mixture (Geometric)			98	107	29	68	90
Mean Fold Improvement Over Mixture (Arithmetic)			120	145	42	121	74

Virus	Clade	Tier	IC ₅₀ (nM)				
			10-1074-40bp-3BNC60	10-1074-60bp-3BNC60	10E8-40bp-3BNC60	10E8-50bp-3BNC60	10E8-60bp-3BNC60
QH-0692	B	2	NT	NT	1.2 (34x)	1.4 (30x)	1.3 (32x)
CAAN5342.A2	B	2	NT	NT	7.3 (>46x)	10 (>32x)	25 (>13x)
THRO4156.18	B	2	24 (2.5x)	19 (3x)	NT	NT	NT
Du172.17	C	2	0.64 (19x)	4.0 (3x)	NT	NT	NT
Du422.1	C	2	NT	NT	2.2 (8x)	3.6 (5x)	5.0 (3x)
CAP45.2.00.G3	C	2	40 (11x)	260 (2x)	11 (2x)	15 (1x)	17 (1x)
CAP210.2.00.E8	C	2	56. (11x)	79 (8x)	NT	NT	NT
T250-4	AG	2	NT	NT	0.92 (53x)	0.66 (74x)	12 (4x)
Geometric Mean IC ₅₀ (nM)			14	35	2.9	3.5	8
Arithmetic Mean IC ₅₀ (nM)			30	90	4.5	6.1	12
Mean Fold Improvement Over Mixture (Geometric)			8.70	3.30	16	13	5.8
Mean Fold Improvement Over Mixture (Arithmetic)			9.3	3.1	20	15	7.6

>100x	50-100x	30-49x	10-29x	2-9x	≤2x	NT
-------	---------	--------	--------	------	-----	----

Figure 4. Synergistic dsDNA-Based Hetero-diFabs

Neutralization of primary HIV-1 strains by hetero-diFabs. IC₅₀s are reported for the hetero-diFabs. See [Tables S1, S2, S3, S4, and S5](#) for IC₅₀s of parental Fabs and IgGs, dsDNA alone, and the non-covalent mixtures of Fabs and dsDNA. As a measure of potential synergy of each hetero-diFab, the molar ratio of the IC₅₀ values for the non-covalent mixture and the hetero-diFab is listed for each strain in parentheses beside the IC₅₀ for the hetero-diFab. NT, not tested. See also [Figure S4](#).



B

Virus	Clade	Tier	IC ₅₀ (nM)			
			TPR12	PG16 Fab	3BNC60 Fab-TPR12	PG16-TPR12-3BNC60
6535.3	B	1B	>1000	750	28	0.9 (30x)
SC422661.8	B	2	600	>1000	1.8	0.1 (15x)
TRO.11	B	2	600	>1000	1.0	0.09 (11x)
CAAN5342.A2	B	2	600	400	10	0.7 (14x)
THRO4156.18	B	2	>1000	910	38	0.6 (59x)
RHPA4259.7	B	2	600	300	0.6	0.05 (11x)
Du172.17	C	2	600	2.7	150	0.1 (27x)
Du422.1	C	2	600	36	360	0.3 (116x)
ZM197M.PB7	C	1B	210	640	11	0.9 (12x)
ZM109F.PB4	C	1B	600	>1000	25	0.4 (59x)
ZM53M.PB12	C	2	600	4.2	7.5	0.02 (209x)
CAP210.2.00.E8	C	2	>1000	9.5	290	0.1 (73x)
Geometric Mean IC ₅₀ (nM)						0.21
Arithmetic Mean IC ₅₀ (nM)						0.37
Mean Fold Improvement Over Best Component (Geometric)						33
Mean Fold Improvement Over Best Component (Arithmetic)						38

>100x	50-100x	30-49x	10-29x	2-9x	≤2x
-------	---------	--------	--------	------	-----

achieve synergy through simultaneous recognition of two different epitopes on the same HIV-1 Env trimer.

To more precisely define optimal intra-epitope separation distances, we evaluated hetero-diFabs with different bridge lengths, finding length-dependent synergy effects. For example, PG16-3BNC60 hetero-diFabs with 40 bp and 50 bp dsDNA bridges showed improved neutralization potencies when compared to the 60 bp (204 Å) version, achieving ≥ 100 -fold potency increases against over half of the tested strains and geometric mean improvements of 98- and 107-fold, respectively (Figure 4 and Table S4). The 40 bp and 50 bp bridges (136 Å and 170 Å, respectively) corresponded to the approximate separation distances between PG16 and 3BNC60 Fabs when bound to the same gp120 within a trimer (147 Å) or to neighboring protomers within open or partially open trimers (167 Å) (Figure S3). In a second length dependency example, 10-1074-40bp-3BNC60 was more potent than 10-1074-60bp-3BNC60

Figure 5. Synergistic Protein-Based Hetero-diFab

(A) Schematic representation of PG16-TPR12-3BNC60 (not to scale). Approximate lengths are indicated (120 Å for the TPR12 linker plus ~ 11 Å for the fused click handles). (B) Neutralization of primary HIV-1 strains. IC₅₀s are reported for PG16-TPR12-3BNC60, the parental components of the reagent (PG16 Fab and 3BNC60 Fab-TPR12), and TPR12 alone. As a measure of potential synergy of PG16-TPR12-3BNC60, the molar ratio of the IC₅₀ values for the most potent component and PG16-TPR12-3BNC60 is listed for each strain in parentheses beside the IC₅₀ for PG16-TPR12-3BNC60. See also Figure S5.

(Figure 4 and Table S4). The ~ 136 Å distance between the two Fabs in 10-1074-40bp-3BNC60 corresponded to the approximate separation between these Fabs bound to the same gp120 (141 Å), whereas 60 bp more closely approximated Fabs bound to neighboring protomers on an open trimer (193 Å) (Figure S3). The 40 bp and 50 bp versions of 10E8-3BNC60 showed consistent synergy (Figure 4 and Table S4); however, the lack of structural information concerning 10E8 binding to Env trimer hindered interpretation of 10E8-containing hetero-diFabs.

A Hetero-diFab Constructed with a Protein Linker Exhibits Synergistic Potency Increases

Bivalent molecules involving dsDNA linkers were effective for demonstrating synergistic neutralization, but a protein reagent would be preferable as an anti-HIV-1 therapeutic. We recently described

a series of protein linkers of various lengths and rigidities (Klein et al., 2014) that can mimic the properties of different lengths of dsDNA. Thus, we can substitute a comparable protein linker for an optimal dsDNA bridge to create a protein reagent capable of simultaneous binding to two different epitopes on a single HIV-1 spike trimer. As a proof-of-principle example, we used sortase-catalyzed protein ligation and click chemistry (Witte et al., 2013) to construct a bivalent reagent analogous to PG16-40bp-3BNC60 by substituting the dsDNA linker with 12 domains of a designed tetratricopeptide-repeat (TPR) protein (Kajander et al., 2007) (Figures 5A and S5). We chose a TPR linker because tandem repeats of TPR domains form a rigid rod-like structure whose length corresponds predictably with the number of repeats, with each domain contributing ~ 10 Å (Kajander et al., 2007). PG16 Fab was expressed with a C-terminal sortase signal, and the C terminus of the 3BNC60 Fab was modified to include 12 TPR repeats and a sortase signal. The tagged Fabs

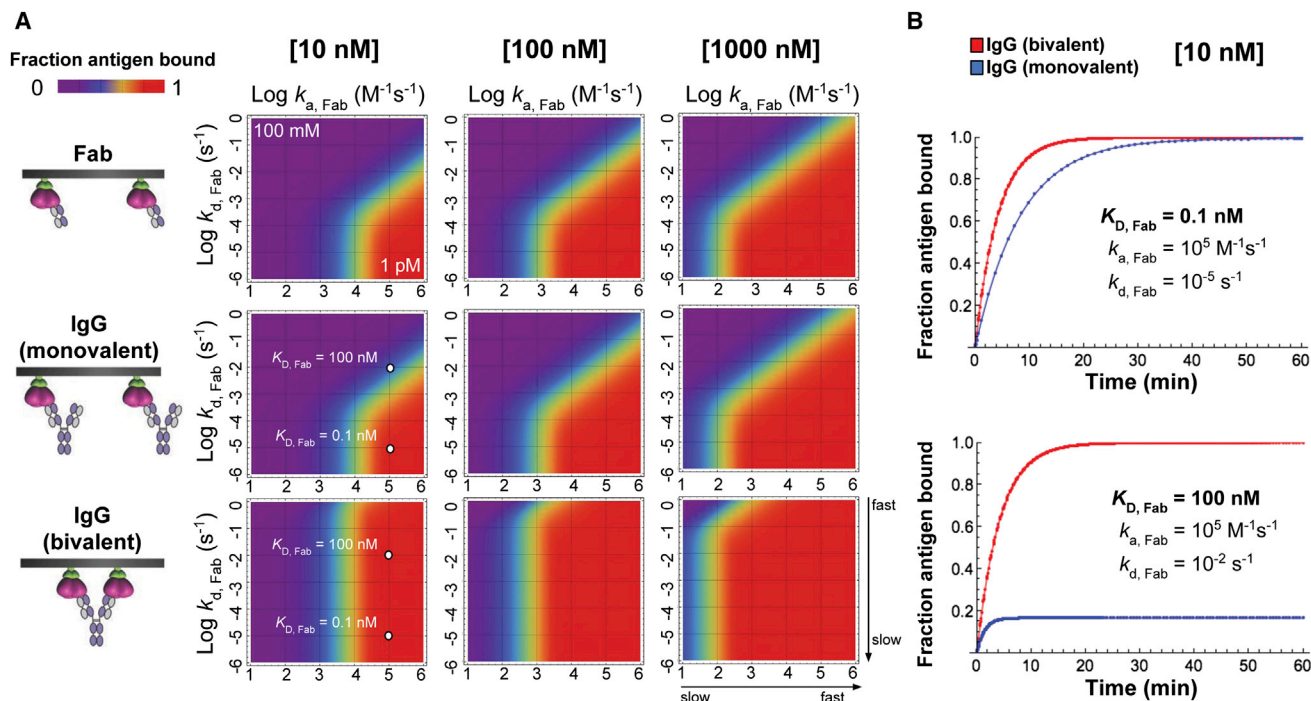


Figure 6. Simulations of Avidity Effects due to Bivalent Binding of IgG to a Tethered Antigen

(A) The fraction of tethered antigen bound by different concentrations of IgG or Fab after 1 hr shown as a heat map (cooler colors representing a lower percentage bound and warmer colors representing a higher percentage bound) as a function of kinetic constants for the IgG-antigen or Fab-antigen interaction. The fraction of antigen bound by a Fab or IgG was calculated as a function of k_a and k_d . The intrinsic affinities are strongest in the lower right corner (1 pM) and weakest in the upper left corner (100 mM) of each graph. For IgG, binding was forced to 100% monovalent binding (middle row) or 100% bivalent binding (bottom row). Saturation by Fabs and IgGs was nearly identical for monovalent binding conditions because the binding kinetics of IgGs would be enhanced by at most 2-fold. Comparisons of the simulations for bivalent binding (bottom row) and monovalent binding (top two panels) showed regions of saturation binding resulting from avidity effects.

(B) The fraction of antigen bound as a function of time for IgGs binding to surface-tethered antigens at an input concentration of 10 nM. When the dissociation rate constant of the Fab portion of the IgG is slow (top) and the input concentration is approximately 100-fold higher than the affinity of the Fab, IgGs can reach saturation binding after an hour whether binding monovalently or bivalently to the surface—hence, avidity effects are not apparent after an hour. However, weakening the affinity of the Fab by making the dissociation rate 1,000-fold faster (bottom) prevents saturation when binding monovalently but has no effect on saturation when binding bivalently—hence, avidity effects are apparent throughout the incubation.

were covalently attached to peptides containing click handles using sortase-catalyzed ligation and then incubated to allow the click reaction to form PG16 Fab linked to 3BNC60 Fab by 12 TPR repeats (PG16-TPR12-3BNC60). Together with the remnants of the click handles, the linker would occupy $\sim 131 \text{ \AA}$, approximately the same length as the dsDNA linker in PG16-40bp-3BNC60 reagent (Figures 5A and S5). The protein-based molecule, PG16-TPR12-3BNC60, exhibited between 11- and >200-fold synergy against 12 primary HIV-1 strains (Figure 5B; 33-fold geometric mean increased potency).

Simulations of the Effects of Avidity on IgG Binding to Tethered Antigens

To better understand the effects of avidity arising from bivalent binding of IgGs to antigens tethered to a surface such as a viral membrane, we used modeling software to simulate the saturation of surface-bound antigens by monovalent Fabs and bivalent IgGs. We chose a 1 hr incubation time based upon conditions under which in vitro neutralization assays are conducted (Montefiori, 2005). We varied the density of the tethered anti-

gens and the concentrations of Fab or IgG and investigated a range of intrinsic association and dissociation rate constants for the binding interaction. The fraction of antigen bound by a Fab or IgG was calculated as a function of on- and off-rates (k_a and k_d), whose ratio (k_d/k_a) is equal to the affinity (K_D , or equilibrium dissociation constant). We compared saturation by Fabs (top row), IgGs in which only monovalent binding was permitted (center row), and IgGs that bound bivalently through crosslinking of neighboring antigens (bottom row) (Figure 6A). As expected, saturation by Fabs and IgGs was nearly identical for monovalent binding conditions (Figure 6A, first two rows). By contrast, across a range of input concentrations, there were k_a and k_d combinations for IgGs binding bivalently that exhibited saturation binding under conditions in which monovalent Fabs and IgGs binding monovalently did not (Figure 6A, bottom row). Thus, consistent with experimental results in the palivizumab/RSV system (Wu et al., 2005), the simulations suggested that bivalency through crosslinking can rescue binding of IgGs whose Fabs exhibit weak binding affinities as a result of fast dissociation rate constants, whereas IgGs whose

Fabs exhibit high affinities because of slow dissociation rates did not display strong avidity enhancement.

The simulations also demonstrate that the effects of avidity on binding are a complicated mixture of kinetics, input concentration, and incubation time. At any particular concentration, the threshold at which avidity is observed is controlled by kinetics rather than affinity because different combinations of kinetic constants yield the same K_D . The kinetic threshold at which avidity effects are observed varies depending on the difference between the input concentration and the K_D . For concentrations near or below the K_D , there is a kinetic threshold such that, for on- and off-rates slower than $\sim 10^3 \text{ M}^{-1}\text{s}^{-1}$ and $\sim 10^{-5} \text{ s}^{-1}$, respectively, avidity enhancement is not observed (Figure 6A). The binding reactions are also affected by the length of incubation, such that the lower the input concentration, the longer it takes to reach saturation (Figure 6B).

We note that the simulations model binding interactions only, whereas our homo- and hetero-diFabs were evaluated for their ability to enhance neutralization of viral infectivity, a process more complicated than binding. For example, neutralization mechanisms may involve conformational changes in Env that were not accounted for in our binding simulation. In addition, kinetics constants for antibody-mediated neutralization of HIV-1 are not known—nor is the fraction of Env spikes on a virion that are required for neutralization or for fusion. In any case, it appears that the kinetic properties of the bNAb Fab components in our reagents were appropriate to realize avidity-enhanced neutralization because hetero-diFab reagents displayed ~ 100 -fold mean improved neutralization potencies. The data therefore support the hypothesis that intra-spike crosslinking by anti-HIV-1 binding molecules represents a valid strategy for increasing potency and resistance to HIV-1 Env mutations.

DISCUSSION

We engineered HIV-1 spike-binding molecules designed to bind bivalently to demonstrate the importance of avidity effects in antibody efficacy in HIV-1 neutralization and to establish that lack of bivalent binding by physiologic IgGs is an additional antibody evasion strategy utilized by HIV-1.

The importance for HIV-1 in maintaining a low spike density to avoid inter-spike crosslinking by IgGs was suggested by the relatively small improvements in neutralization potencies of intact anti-HIV-1 IgGs compared with their Fab counterparts (Klein and Bjorkman, 2010) and by the discovery that polyreactivity increased the apparent affinity of anti-HIV-1 antibodies through a mechanism of heterologation (Mouquet et al., 2010). Comparison of the neutralization potencies of IgGs versus Fabs in the current study provides further support for the observation that anti-HIV-1 IgGs generally exhibit relatively small increased potencies compared to Fabs. To quantify potential avidity effects, we previously defined the molar neutralization ratio (MNR) for IgG versus Fab forms of an antibody as $\text{IC}_{50} \text{ Fab (nM)}/\text{IC}_{50} \text{ IgG (nM)}$. In the absence of avidity or other advantages of the IgG compared with the Fab (e.g., increased size), the ratio would be 2.0 (Klein and Bjorkman, 2010). In the current study, the mean MNR for PG16, an IgG that cannot exhibit intra-spike

crosslinking (Julien et al., 2013b), was 8.0 (data from Figure 2C), similar to the 10.5 mean MNR in a previous study (West et al., 2012). These values are lower than MNRs observed for IgGs against densely packed viruses, which can be over 1,000, but are consistent with a limited amount of inter-spike crosslinking by anti-HIV-1 IgGs whose epitopes on neighboring spikes are accessible to simultaneous engagement of the combining sites of the two Fabs of an IgG, which are separated by $\sim 150 \text{ \AA}$ (Klein and Bjorkman, 2010). Our current results suggested that inter-spike crosslinking can be increased by creating homo-diFabs with 70 bp–100 bp dsDNA linkers (Figures 2B and S2D). These linkers would separate the combining sites of the Fabs by ≥ 240 – 340 \AA , distances that should enhance inter-spike crosslinking.

Indirect evidence for the hypothesis that HIV-1 evolved a low spike density to avoid inter-spike crosslinking IgGs comes from studies of a cytoplasmic tail deletion in the simian immunodeficiency virus (SIV) spike trimer. Cytoplasmic tail deletion has been suggested to increase the number of spikes per virion (Zingler and Littman, 1993) and/or the spike mobility in the virion bilayer (Crooks et al., 2008), both of which could enhance inter-spike crosslinking. Although tail-deleted mutant viruses can be produced in vitro, propagation of the virus in macaques favors viruses containing the full-length envelope spike (Zingler and Littman, 1993). These findings are consistent with the idea that an intact host immune system selects against those viruses that facilitate the ability of host IgGs to bind bivalently through inter-spike crosslinking.

Here, we present a method to create potential intra-spike crosslinking antibody-based molecules using dsDNA- and protein-based linkers and demonstrate that these reagents can exhibit up to three orders of magnitude increases in neutralization potency. We argue that the optimized versions of our engineered molecules achieve potency increases through intra-spike, rather than inter-spike, crosslinking because (1) distances measured between epitopes on virion-bound spike trimers corresponded to approximate intra-epitope distances on HIV-1 spike trimer structures, and (2) increases in inter-spike crosslinking by homotypic and heterotypic reagents should not exhibit sharp linker length-dependent neutralization potencies because distances between spikes vary within a single virion and between virions. The latter point is valid even if HIV-1 spikes are clustered on mature virions, as suggested by fluorescence nanoscopy (Chojnacki et al., 2012), but not cryoelectron microscopy (Liu et al., 2008; Zhu et al., 2006). Whether HIV-1 spikes cluster upon encountering a target cell to form an entry claw (Sougrat et al., 2007) is not relevant to the mechanism of action of our reagents because neutralization assays are conducted by incubating potential inhibitors with virions prior to addition of target cells (Montefiori, 2005), a mechanism that is also presumably relevant for most in vivo interactions of antibody and antibody-like inhibitors. Because avidity effects require recognition of two or more antigens tethered to the same surface, another potential action of our reagents, inter-virion crosslinking, would not result in avidity effects by analogy to the lack of avidity enhancement for an IgG binding two soluble antigens, one per Fab. In this respect, we note that, although IgAs are capable of inter-virion crosslinking (Stieh et al., 2014), conversion of IgG bNAbs to

IgAs did not result in potency increases (Kunert et al., 2004; Wolbank et al., 2003).

The use of dsDNA- and protein-based molecular rules to measure inter-epitope distances presented here can be used to probe conformations of virion-bound Env trimers. By contrast, EM and X-ray structures (Bartesaghi et al., 2013; Julien et al., 2013a; Lyumkis et al., 2013; Pancera et al., 2014) cannot capture dynamic information concerning Env conformations during neutralization. Single-molecule fluorescence resonance energy transfer (smFRET) measurements suggested that Env trimers on the surface of HIV-1 virions transition between different conformations (Munro et al., 2014), and spike trimers have been visualized by EM in different conformations: the closed structure of unliganded trimers and trimers associated with VRC01-like bNAbs (Bartesaghi et al., 2013; Liu et al., 2008; Lyumkis et al., 2013) (also observed in Fab-bound crystal structures [Julien et al., 2013a; Pancera et al., 2014]), a CD4- and/or 17b-bound open structure (Liu et al., 2008; Tran et al., 2012), and a partially open b12-bound structure (Liu et al., 2008) (Figure 3A). Homo- and hetero-diFabs joined by different lengths of dsDNA bridges offer a new methodology to probe Env trimer conformational states on virions and potentially to address strain-specific conformational differences.

Homo-diFabs constructed from VRC01-like bNAbs showed greatest potency when binding to epitopes separated by distances most closely approximating the open structure (Liu et al., 2008; Tran et al., 2012), rather than the closed structure observed for soluble and virion-associated spike trimers bound to VRC01-like Fabs (Bartesaghi et al., 2013; Liu et al., 2008; Lyumkis et al., 2013) (Figures 2B, 3, S3, and S4). These results suggest that optimal intra-spike crosslinking molecules can inhibit a different state than recognized by monovalent Fabs binding to spike trimers in static EM and X-ray structures (Bartesaghi et al., 2013; Liu et al., 2008; Lyumkis et al., 2013; Merk and Subramaniam, 2013; Tran et al., 2012). If so, one Fab of a homo-diFab could first bind to its epitope on a closed trimer, allowing the second Fab to latch on to a transiently populated open form of that trimer. Alternatively, binding of the first Fab may trap the trimer into a conformation allowing increased accessibility of the second Fab, or both Fabs could bind simultaneously to a transiently appearing open trimer. Interestingly, the distance dependence of two CD4bs antibodies, 3BNC60 and b12, was more strongly pronounced for a Tier 1B HIV-1 strain, 6535.3, than for Tier 2 or 3 strains against which the homo-diFabs were tested (Figures 2B and S2). Tier categorization of HIV-1 strains refers to the sensitivity of a strain to antibody neutralization, with Tier 1 strains being more sensitive in general to antibodies than Tier 2 or 3 strains (Seaman et al., 2010). The differences in length dependence for CD4bs homo-diFabs may reflect differences in conformational variability within Env trimers from different tiers, with Tier 1 Env perhaps more easily able to adopt the open conformations likely recognized by the CD4bs antibodies with optimal bridge lengths.

For the PG16-3BNC60 hetero-diFabs, the optimal 40 bp and 50 bp bridge lengths (136 Å and 170 Å, respectively) corresponded to the approximate separation distances between PG16 and 3BNC60 Fabs when bound to the same gp120 within a trimer (147 Å) or to neighboring protomers within open or

partially open trimers (167 Å) (Figure S3). In a second hetero-diFab bridge length dependency example, 10-1074-40bp-3BNC60 was more potent than 10-1074-60bp-3BNC60 (Figure 4 and Table S4). The ~136 Å distance between the two Fabs in 10-1074-40bp-3BNC60 corresponded to the approximate separation between these Fabs bound to the same gp120 (141 Å), whereas 60 bp more closely approximated Fabs bound to neighboring protomers on an open trimer (193 Å) (Figure S3). In general, it is more difficult to deduce information about Env trimer conformations recognized by hetero-diFabs because the intra-epitope distance is the same in the three conformations for Fabs binding to the same gp120 subunit within an Env trimer (Figure S3), and length-dependence data for some of the hetero-diFabs (e.g., 10-1074-40bp-3BNC60) were consistent with binding to a single gp120 within an Env trimer, as well as to adjacent gp120s (Figure S2). However, whether binding to the same or to adjacent protomers within the spike trimer, the increased synergy of optimal hetero-diFabs suggested a mechanism in which the more potent/tighter-binding Fab of the hetero-diFab initially bound to the viral spike, thereby allowing the second Fab, even when only weakly neutralizing on its own, to attach.

In summary, our results demonstrated that optimal length homo- and hetero-diFabs are capable of synergistic effects that increased neutralization potencies and, in some cases, allowed neutralization of viral strains resistant to conventional IgGs. These results are consistent with the hypothesis that most anti-HIV-1 IgGs bind monovalently to single Env spikes, which leaves them vulnerable to Env mutations that weaken monovalent interactions but would still permit bivalent interactions (Klein and Bjorkman, 2010). The demonstration that anti-HIV-1 reagents designed to be capable of intra-spike binding with avidity can more potently and broadly neutralize HIV-1 than conventional anti-spike IgGs is relevant to the choice of anti-HIV-1 proteins or genes to be delivered passively to prevent infection or suppress active infections. Bi-specific antibodies that simultaneously bind to HIV-1 Env and to CD4 or CCR5 host receptors on the target cell represent a conceptually distinct method to increase the potency and breadth of anti-HIV-1 reagents (Pace et al., 2013). In contrast to these reagents, antibodies that achieve synergy via bivalent binding to Env by intra-spike crosslinking offer significant advantages for passive delivery; for example, neutralizing antibodies against HIV-1 Env protect more effectively in vivo than antibodies against CD4 (Pegu et al., 2014), and anti-self antibodies such as anti-CD4 IgGs have short half-lives in vivo (Bruno and Jacobson, 2010). We propose that the ideal therapeutic molecule would utilize avidity achieved by intra-spike crosslinking to reduce the concentration required for sterilizing immunity and render the low spike density of HIV-1 irrelevant to its efficacy. Moreover, analogous to using several drugs or antibodies during anti-retroviral therapy, simultaneous binding to different HIV-1 epitopes should reduce or abrogate sensitivity to Env mutations.

EXPERIMENTAL PROCEDURES

Expression and Purification of Fabs

Genes encoding IgG light-chain genes were modified by site-directed mutagenesis to replace Cys263_{Light Chain}, the C-terminal cysteine that forms a

disulfide bond with Cys233_{heavy chain}, with a serine. Modified light-chain genes and genes encoding 6x-His- or StrepII-tagged Fab heavy chains (V_H-C_H1-tag) were subcloned separately into the pTT5 mammalian expression vector (NRC Biotechnology Research Institute). Fabs were expressed by transient transfection in HEK293-6E (NRC Biotechnology Research Institute) cells as described (Diskin et al., 2011) and purified from supernatants by Ni-NTA or StrepII affinity chromatography followed by size exclusion chromatography in PBS pH 7.4 using a Superdex 200 10/300 or Superdex 200 16/600 column (Amersham Biosciences).

IgG heterodimers were expressed and purified as described in the [Extended Experimental Procedures](#).

DNA Conjugation to Fabs

DNA was conjugated to free thiol-containing Fabs using a modified version of a previously described protocol (Hendrickson et al., 1995). Briefly, Fabs were reduced in a buffer containing 10 mM TCEP-HCl (pH 7–8) for 2 hr and then buffer exchanged three times over Zeba desalting columns (Thermo Scientific). The percentage of reduced Fab was determined using Invitrogen's Measure-IT Thiol Assay. Concurrently, a 5–20 base ssDNA containing a 5' amino group (Integrated DNA Technologies, IDT-DNA) was incubated with a 100-fold molar excess of an amine-to-sulphydryl crosslinker (Sulfo-SMCC; Thermo Scientific) for 30 min to form a maleimide-activated DNA strand, which was buffer exchanged as described above. The reduced Fab and activated ssDNA were incubated overnight, and the Fab-ssDNA conjugate was purified by Ni-NTA or StrepII affinity chromatography (GE Biosciences) to remove unreacted Fab and ssDNA.

ssDNA was synthesized, phosphorylated, and PAGE purified by Integrated DNA Technologies. For di-Fabs containing dsDNA bridges longer than 40 bp, complementary ssDNAs were annealed by heating (95°C) and cooling (room temperature) to create dsDNA containing overhangs complementary to the Fab-ssDNA conjugates. dsDNA was purified by size exclusion chromatography (Superdex 200 10/300) and incubated overnight with the corresponding tagged Fab-ssDNA conjugates. Homo- and hetero-diFab reagents were purified by Ni-NTA and StrepII affinity chromatography when appropriate to remove free DNA and excess Fab-ssDNA conjugates, treated with T4 DNA ligase (New England Biolabs), and purified again by size exclusion chromatography (Figure S1B). To make di-Fabs containing dsDNA bridge lengths less than 40 bp, two complementary ssDNA-conjugated Fabs were incubated at 37°C without a dsDNA bridge and then purified as described above. Protein-DNA reagents were stable at 4°C for >6 months as assessed by SDS-PAGE.

Bridge and linker sequences are listed in [Extended Experimental Procedures](#).

Characterization of DNA-Fab Reagents

Fractions from the center of an SEC elution peak were concentrated using Amicon Ultra-15 Centrifugal Filter Units (Millipore) (MW cutoff = 10 kDa) to a volume of 500 μ l, and DLS measurements were performed on a DynaPro NanoStar (Wyatt Technology) using the manufacturer's suggested settings. Hydrodynamic radii were determined as described (Dev and Surolia, 2006). Briefly, a nonlinear least-squares fitting algorithm was used to fit the measured correlation function to obtain a decay rate. The decay rate was converted to the diffusion constant that can be interpreted as the hydrodynamic radius via the Stokes-Einstein equation.

Hetero-diFab with TPR Linker

PG16-TPR12-3BNC60, a C-to-C linked hetero-diFab containing 12 consensus TPR domains (Kajander et al., 2007) as a protein linker (Klein et al., 2014), was prepared from modified PG16 and 3BNC60 Fabs using a combination of sortase-catalyzed peptide ligation and click chemistry (Witte et al., 2013). The C terminus of the PG16 Fab heavy chain was modified to include the amino acid sequence GGGGASLPETGGLNIDFEAQKIEWHEHHHHH, comprising a flexible linker, the recognition sequence for *S. aureus* Sortase A (underlined), a BirA tag, and a 6x-His tag. The C terminus of the 3BNC60 Fab heavy-chain C terminus was modified to include a (Gly₄Ser)₃ linker followed by 12 tandem TPR domains and the amino acid sequence ASGGGSGGGGSGGGGS LPETGGHHHHHH, comprising a second (Gly₄Ser)₃ linker, the Sortase A

recognition sequence (underlined), and a 6x-His tag. The Fabs were expressed in HEK-6E cells and purified with Ni-NTA and gel filtration chromatography as described above. Peptides (GGGK with C-terminal azide and cyclooctyne click handles) were synthesized by GenScript, and sortase-catalyzed peptide ligation was used to attach the azide-containing peptide to PG16 Fab and the cyclooctyne-containing peptide to the 3BNC60-TPR12 fusion protein as described (Guimaraes et al., 2013). Approximate yields after each sortase reaction were ~30%. Peptide-ligated PG16 and 3BNC60 Fabs were passed over a Ni-NTA column to remove His-tagged enzyme and Fabs that did not lose their His tags during the reaction, mixed at equimolar ratios, and the click reaction was accomplished by incubating overnight at 25°C. The approximate yield for the click reaction was ~65%. The resulting PG16-TPR12-3BNC60 hetero-diFab was purified by size exclusion chromatography to remove unreacted Fabs for an overall yield of ~22%.

Measurements of Intra-Spike Distances

To derive predicted distances between two adjacent Fab bound to HIV-1 Env, we superimposed Fabs bound to their epitopes on the structures of Env trimers in three different conformations: closed (a 4.7 Å crystal structure of a gp140 SOSIP trimer; PDB code 4NCO), open (a 9 Å EM structure of a SOSIP trimer-17b Fab complex (Tran et al., 2012); coordinates obtained from S. Subramaniam), and partially open (an ~20 Å EM structure of a viral spike bound to b12 Fab; PDB code 3DNL). The positions of the C_H1 and C_L domains in Fab structures used for docking were adjusted to create Fabs with the average elbow bend angle found in a survey of human Fab structures (Stanfield et al., 2006). The V_H-V_L domains of the adjusted Fabs were then superimposed on crystal structures of Fab-gp120 or Fab-gp140 complexes (PDB codes 3NGB, 2NY7, and 4CNO for complexes with VRC01, b12, and PGT122 Fabs, respectively) or a PG16-epitope scaffold complex (PDB code 4DQO). The position on Env trimer of 10-1074, a clonal variant of the PGT121-PGT123 family (Mouquet et al., 2012), was approximated using the 4CNO gp140-PGT122 structure. In other cases, related antibodies, e.g., PG9/PG16 and VRC01/3BNC117/3BNC60, were also assumed to bind similarly. The complex structures were superimposed on the Env trimer structures by aligning the common portions. The distance between the Cys233_{heavy chain} carbon- α atoms of adjacent Fabs was then measured using PyMol (Schrodinger, 2011) to approximate the length of dsDNA bridges attached to Cys233_{heavy chain}. Measurements derived using other EM structures for the closed and open trimers (PDB codes 3DNN, 3J5M, and 3DNO) or using a recent 3.5 Å Env trimer crystal structure (Pancera et al., 2014) resulted in differences of ≤ 10 Å for analogous distance measurements.

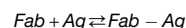
In Vitro Neutralization Assays

Neutralization of pseudoviruses derived from primary HIV-1 isolates was monitored by the reduction of HIV-1 Tat-induced luciferase reporter gene expression in the presence of a single round of pseudovirus infection in TZM-bl cells as described previously (Montefiori, 2005) and in the [Extended Experimental Procedures](#).

Simulation of Fab and IgG Saturation of Surface-Bound Antigens

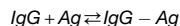
Numerical analysis (Mathematica, v. 10) was used to simulate saturation of surface-bound antigens by monovalent Fabs (Equation 1), bivalent IgGs to unpaired antigen (Ag) (Equation 2), and paired antigen (pAg) (Equations 3 and 4), where "paired antigen" was defined as antigens that are spaced such that an IgG can bind two epitopes simultaneously (e.g., intra-spike crosslinking of two epitopes on the same viral spike or inter-spike crosslinking between two viral spikes). In the bivalent model (Equations 3 and 4), the surface concentrations of antigen and IgG-antigen complexes were approximated by the inverse of the volume of a sphere (V_s) with radius equal to the hydrodynamic radius of the molecule multiplied by Avogadro's number (N_a) as described previously (Müller et al., 1998).

Fab binding to antigen:



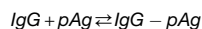
$$\frac{d[Fab - Ag]}{dt} = k_a[Fab][Ag] - k_d[Fab - Ag] \quad (\text{Equation 1})$$

IgG binding to unpaired antigen:



$$\frac{d[\text{IgG} - \text{Ag}]}{dt} = 2k_a[\text{IgG}][\text{Ag}] - k_d[\text{IgG} - \text{Ag}] \quad (\text{Equation 2})$$

IgG binding to paired antigen:



$$\frac{d[\text{IgG} - \text{pAg}]}{dt} = 2k_a[\text{IgG}][\text{pAg}] - k_d[\text{IgG} - \text{pAg}] - \frac{d[\text{IgG} - \text{pAg}_2]}{dt} \quad (\text{Equation 3})$$

$$\frac{d[\text{IgG} - \text{pAg}_2]}{dt} = k_a[\text{IgG} - \text{pAg}] \frac{1}{V_s N_a} [\text{pAg}] \frac{1}{V_s N_a} - 2k_d[\text{IgG} - \text{pAg}_2] \quad (\text{Equation 4})$$

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, five figures, and six tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.01.016>.

AUTHOR CONTRIBUTIONS

R.P.G. and P.J.B. conceived the study; J.S.K. performed simulations to assess avidity effects; R.P.G. and M.S.P. prepared dsDNA-Fab reagents; R.P.G., M.S.P., J.S.K., S.B., and A.P.W. perfected methods to attach dsDNA to Fabs; R.P.G. and M.S.S. performed neutralization assays; R.P.G., A.P.W., and P.J.B. analyzed the data; and R.P.G., M.C.N. and P.J.B. wrote the paper with contributions from all co-authors.

ACKNOWLEDGMENTS

We thank Martin Witte, Jessica Ingram, Chris Theile, and Hidde Ploegh for advice and help with sortase/click chemistry experiments; Sriram Subramaniam for helpful discussions, coordinates, and files to make schematic figures; David Baltimore, Bil Clemons, Ron Diskin, Jennifer Keefe, Sarkis Mazmanian, Stuart Sievers, Louise Scharf, and Kai Zinn for suggestions; Jost Vielmetter and the Caltech Protein Expression Center for assistance with protein production and neutralization assay development; Priyanthi Gnanaprasam for doing in-house neutralization assays; the NIH AIDS Reagent Program, the Fraunhofer Institut IBMT, and Rene Mares for pseudoviruses; Erin Isaza, Siduo Jiang, and Jennifer Keefe for reagents; and Sonal N. Patel and Marta Murphy for help with figures. This work was supported by the Director's Pioneer Award (1DP1OD006961-01 to P.J.B.), National Institutes of Health HIVRAD (P01 AI100148 to P.J.B.), and Collaboration for AIDS Vaccine Discovery (CAVD) grants with support from the Bill and Melinda Gates Foundation (grant 38660 [P.J.B.] and grant 1032144 [M.S.S.]). P.J.B. and M.C.N. are HHMI Investigators.

Received: October 16, 2014

Revised: November 21, 2014

Accepted: December 16, 2014

Published: January 29, 2015

REFERENCES

Bartesaghi, A., Merk, A., Borgnia, M.J., Milne, J.L., and Subramaniam, S. (2013). Prefusion structure of trimeric HIV-1 envelope glycoprotein determined by cryo-electron microscopy. *Nat. Struct. Mol. Biol.* 20, 1352–1357.

Bednar, J., Furrer, P., Katritch, V., Stasiak, A.Z., Dubochet, J., and Stasiak, A. (1995). Determination of DNA persistence length by cryo-electron microscopy.

Separation of the static and dynamic contributions to the apparent persistence length of DNA. *J. Mol. Biol.* 254, 579–594.

Bhatia, A.K., Kaushik, R., Campbell, N.A., Pontow, S.E., and Ratner, L. (2009). Mutation of critical serine residues in HIV-1 matrix result in an envelope incorporation defect which can be rescued by truncation of the gp41 cytoplasmic tail. *Virology* 384, 233–241.

Bruno, C.J., and Jacobson, J.M. (2010). Ibalizumab: an anti-CD4 monoclonal antibody for the treatment of HIV-1 infection. *J. Antimicrob. Chemother.* 65, 1839–1841.

Chertova, E., Bess, J.W., Jr., Crise, B.J., Sowder, R.C., II, Schaden, T.M., Hilburn, J.M., Hoxie, J.A., Benveniste, R.E., Lifson, J.D., Henderson, L.E., and Arthur, L.O. (2002). Envelope glycoprotein incorporation, not shedding of surface envelope glycoprotein (gp120/SU), is the primary determinant of SU content of purified human immunodeficiency virus type 1 and simian immunodeficiency virus. *J. Virol.* 76, 5315–5325.

Chi, Q., Wang, G., and Jiang, J. (2013). The persistence length and length per base of single-stranded DNA obtained from fluorescence correlation spectroscopy measurements using mean field theory. *Physica A: Stat. Mech. Appl.* 392, 1072–1079.

Chojnacki, J., Staudt, T., Glass, B., Bingen, P., Engelhardt, J., Anders, M., Schneider, J., Müller, B., Hell, S.W., and Kräusslich, H.G. (2012). Maturation-dependent HIV-1 surface protein redistribution revealed by fluorescence nanoscopy. *Science* 338, 524–528.

Crooks, E.T., Jiang, P., Franti, M., Wong, S., Zwick, M.B., Hoxie, J.A., Robinson, J.E., Moore, P.L., and Binley, J.M. (2008). Relationship of HIV-1 and SIV envelope glycoprotein trimer occupation and neutralization. *Virology* 377, 364–378.

Dev, S., and Suroli, A. (2006). Dynamic light scattering study of peanut agglutinin: size, shape and urea denaturation. *J. Biosci.* 31, 551–556.

Diskin, R., Scheid, J.F., Marcovecchio, P.M., West, A.P., Jr., Klein, F., Gao, H., Gnanaprasam, P.N., Abadir, A., Seaman, M.S., Nussenzweig, M.C., and Bjorkman, P.J. (2011). Increasing the potency and breadth of an HIV antibody by using structure-based rational design. *Science* 334, 1289–1293.

Guimaraes, C.P., Witte, M.D., Theile, C.S., Bozkurt, G., Kundrat, L., Blom, A.E., and Ploegh, H.L. (2013). Site-specific C-terminal and internal loop labeling of proteins using sortase-mediated reactions. *Nat. Protoc.* 8, 1787–1799.

Hendrickson, E.R., Truby, T.M., Joerger, R.D., Majarian, W.R., and Ebersole, R.C. (1995). High sensitivity multianalyte immunoassay using covalent DNA-labeled antibodies and polymerase chain reaction. *Nucleic Acids Res.* 23, 522–529.

Huang, J., Ofek, G., Laub, L., Louder, M.K., Doria-Rose, N.A., Longo, N.S., Imamichi, H., Bailer, R.T., Chakrabarti, B., Sharma, S.K., et al. (2012). Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. *Nature* 491, 406–412.

Icenogle, J., Shiwen, H., Duke, G., Gilbert, S., Rueckert, R., and Andereg, J. (1983). Neutralization of poliovirus by a monoclonal antibody: kinetics and stoichiometry. *Virology* 127, 412–425.

Julien, J.P., Cupo, A., Sok, D., Stanfield, R.L., Lyumkis, D., Deller, M.C., Klasse, P.J., Burton, D.R., Sanders, R.W., Moore, J.P., et al. (2013a). Crystal structure of a soluble cleaved HIV-1 envelope trimer. *Science* 342, 1477–1483.

Julien, J.P., Lee, J.H., Cupo, A., Murin, C.D., Derking, R., Hoffenberg, S., Caulfield, M.J., King, C.R., Marozsan, A.J., Klasse, P.J., et al. (2013b). Asymmetric recognition of the HIV-1 trimer by broadly neutralizing antibody PG9. *Proc. Natl. Acad. Sci. USA* 110, 4351–4356.

Kajander, T., Cortajarena, A.L., Mochrie, S., and Regan, L. (2007). Structure and stability of designed TPR protein superhelices: unusual crystal packing and implications for natural TPR proteins. *Acta Crystallogr. D Biol. Crystallogr.* 63, 800–811.

Klein, J.S., and Bjorkman, P.J. (2010). Few and far between: how HIV may be evading antibody avidity. *PLoS Pathog.* 6, e1000908.

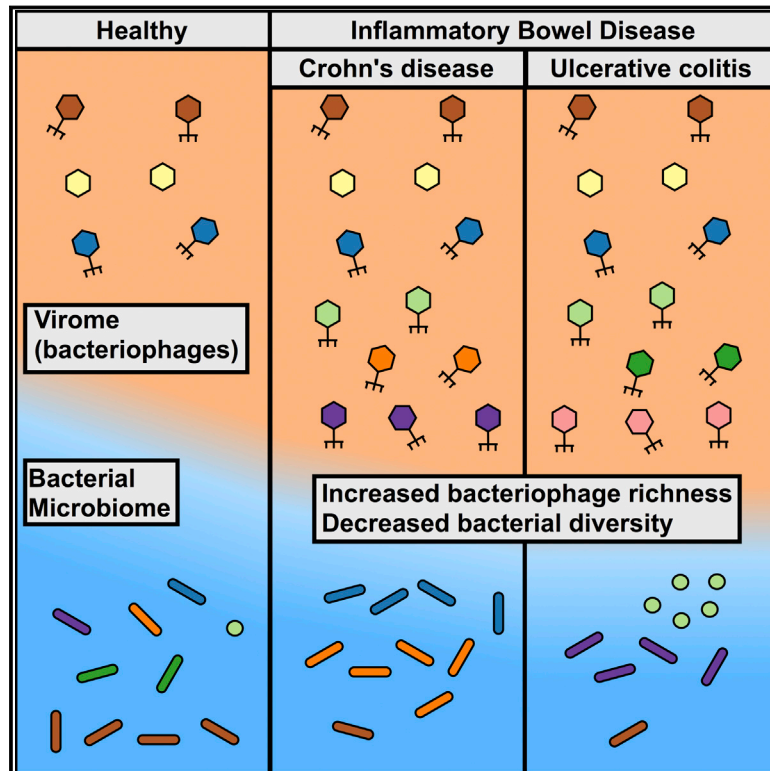
Klein, J.S., Gnanaprasam, P.N., Galimidi, R.P., Foglesong, C.P., West, A.P., Jr., and Bjorkman, P.J. (2009). Examination of the contributions of size and

- avidity to the neutralization mechanisms of the anti-HIV antibodies b12 and 4E10. *Proc. Natl. Acad. Sci. USA* 106, 7385–7390.
- Klein, F., Mouquet, H., Dosenovic, P., Scheid, J.F., Scharf, L., and Nussenzweig, M.C. (2013). Antibodies in HIV-1 vaccine development and therapy. *Science* 341, 1199–1204.
- Klein, J.S., Jiang, S., Galimidi, R.P., Keffe, J.R., and Bjorkman, P.J. (2014). Design and characterization of structured protein linkers with differing flexibilities. *Protein Eng. Des. Sel.* 27, 325–330.
- Kunert, R., Wolbank, S., Stiegler, G., Weik, R., and Kattinger, H. (2004). Characterization of molecular features, antigen-binding, and in vitro properties of IgG and IgM variants of 4E10, an anti-HIV type 1 neutralizing monoclonal antibody. *AIDS Res. Hum. Retroviruses* 20, 755–762.
- Liljeroos, L., Krzyzaniak, M.A., Helenius, A., and Butcher, S.J. (2013). Architecture of respiratory syncytial virus revealed by electron cryotomography. *Proc. Natl. Acad. Sci. USA* 110, 11133–11138.
- Liu, J., Bartesaghi, A., Borgnia, M.J., Sapiro, G., and Subramaniam, S. (2008). Molecular architecture of native HIV-1 gp120 trimers. *Nature* 455, 109–113.
- Luftig, M.A., Mattu, M., Di Giovine, P., Geleziunas, R., Hrin, R., Barbato, G., Bianchi, E., Miller, M.D., Pessi, A., and Carfi, A. (2006). Structural basis for HIV-1 neutralization by a gp41 fusion intermediate-directed antibody. *Nat. Struct. Mol. Biol.* 13, 740–747.
- Lyumkis, D., Julien, J.P., de Val, N., Cupo, A., Potter, C.S., Klasse, P.J., Burton, D.R., Sanders, R.W., Moore, J.P., Carragher, B., et al. (2013). Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. *Science* 342, 1484–1490.
- Mattes, M.J. (2005). Binding parameters of antibodies: pseudo-affinity and other misconceptions. *Cancer Immunol. Immunother.* 54, 513–516.
- Merk, A., and Subramaniam, S. (2013). HIV-1 envelope glycoprotein structure. *Curr. Opin. Struct. Biol.* 23, 268–276.
- Montefiori, D.C. (2005). Evaluating neutralizing antibodies against HIV, SIV, and SHIV in luciferase reporter gene assays. In *Curr. Protoc. Immunol.*, John E. Coligan, ed. (John Wiley & Sons), pp. 64.12.11:12.11.1–12.11.17.
- Mouquet, H., Scheid, J.F., Zoller, M.J., Krogsgaard, M., Ott, R.G., Shukair, S., Artyomov, M.N., Pietzsch, J., Connors, M., Pereyra, F., et al. (2010). Polyreactivity increases the apparent affinity of anti-HIV antibodies by heterologation. *Nature* 467, 591–595.
- Mouquet, H., Scharf, L., Euler, Z., Liu, Y., Eden, C., Scheid, J.F., Halper-Stromberg, A., Gnanapragasam, P.N., Spencer, D.I., Seaman, M.S., et al. (2012). Complex-type N-glycan recognition by potent broadly neutralizing HIV antibodies. *Proc. Natl. Acad. Sci. USA* 109, E3268–E3277.
- Müller, K.M., Arndt, K.M., Strittmatter, W., and Plückthun, A. (1998). The first constant domain (CH1) and C(L) of an antibody used as heterodimerization domain for bispecific miniantibodies. *FEBS Lett.* 422, 259–264.
- Munro, J.B., Gorman, J., Ma, X., Zhou, Z., Arthos, J., Burton, D.R., Koff, W.C., Courtney, J.R., Smith, A.B., 3rd, Kwong, P.D., et al. (2014). Conformational dynamics of single HIV-1 envelope trimers on the surface of native virions. *Science* 346, 759–763.
- Pace, C.S., Song, R., Ochsenbauer, C., Andrews, C.D., Franco, D., Yu, J., Oren, D.A., Seaman, M.S., and Ho, D.D. (2013). Bispecific antibodies directed to CD4 domain 2 and HIV envelope exhibit exceptional breadth and picomolar potency against HIV-1. *Proc. Natl. Acad. Sci. USA* 110, 13540–13545.
- Pancera, M., Zhou, T., Druz, A., Georgiev, I.S., Soto, C., Gorman, J., Huang, J., Acharya, P., Chuang, G.Y., Ofek, G., et al. (2014). Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature* 514, 455–461.
- Pegu, A., Yang, Z.Y., Boyington, J.C., Wu, L., Ko, S.Y., Schmidt, S.D., McKee, K., Kong, W.P., Shi, W., Chen, X., et al. (2014). Neutralizing antibodies to HIV-1 envelope protect more effectively in vivo than those to the CD4 receptor. *Sci. Transl. Med.* 6, 243ra288.
- Roben, P., Moore, J.P., Thali, M., Sodroski, J., Barbas, C.F., 3rd, and Burton, D.R. (1994). Recognition properties of a panel of human recombinant Fab fragments to the CD4 binding site of gp120 that show differing abilities to neutralize human immunodeficiency virus type 1. *J. Virol.* 68, 4821–4828.
- Schaefer, W., Regula, J.T., Böhner, M., Schanzer, J., Croasdale, R., Dürr, H., Gassner, C., Georges, G., Kettenberger, H., Imhof-Jung, S., et al. (2011). Immunoglobulin domain crossover as a generic approach for the production of bispecific IgG antibodies. *Proc. Natl. Acad. Sci. USA* 108, 11187–11192.
- Scheid, J.F., Mouquet, H., Ueberheide, B., Diskin, R., Klein, F., Oliveira, T.Y., Pietzsch, J., Fenyo, D., Abadir, A., Velinzon, K., et al. (2011). Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science* 333, 1633–1637.
- Schofield, D.J., Stephenson, J.R., and Dimmock, N.J. (1997). Variations in the neutralizing and haemagglutination-inhibiting activities of five influenza A virus-specific IgGs and their antibody fragments. *J. Gen. Virol.* 78, 2431–2439.
- Schrödinger, L. (2011). The PyMOL Molecular Graphics System (The PyMOL Molecular Graphics System).
- Seaman, M.S., Janes, H., Hawkins, N., Grandpre, L.E., Devoy, C., Giri, A., Coffey, R.T., Harris, L., Wood, B., Daniels, M.G., et al. (2010). Tiered categorization of a diverse panel of HIV-1 Env pseudoviruses for assessment of neutralizing antibodies. *J. Virol.* 84, 1439–1452.
- Sougrat, R., Bartesaghi, A., Lifson, J.D., Bennett, A.E., Bess, J.W., Zabransky, D.J., and Subramaniam, S. (2007). Electron tomography of the contact between T cells and SIV/HIV-1: implications for viral entry. *PLoS Pathog.* 3, e63.
- Stanfield, R.L., Zemla, A., Wilson, I.A., and Rupp, B. (2006). Antibody elbow angles are influenced by their light chain class. *J. Mol. Biol.* 357, 1566–1574.
- Stieh, D.J., King, D.F., Klein, K., Liu, P., Shen, X., Hwang, K., Ferrari, G., Montefiori, D.C., Haynes, B., Pitisuttithum, P., et al. (2014). Aggregate complexes of HIV-1 induced by multimeric antibodies. *Retrovirology* 11, 78.
- Tran, E.E., Borgnia, M.J., Kuybeda, O., Schauder, D.M., Bartesaghi, A., Frank, G.A., Sapiro, G., Milne, J.L., and Subramaniam, S. (2012). Structural mechanism of trimeric HIV-1 envelope glycoprotein activation. *PLoS Pathog.* 8, e1002797.
- Walker, L.M., Phogat, S.K., Chan-Hui, P.Y., Wagner, D., Phung, P., Goss, J.L., Wrin, T., Simek, M.D., Fling, S., Mitcham, J.L., et al.; Protocol G Principal Investigators (2009). Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science* 326, 285–289.
- West, A.P., Jr., Galimidi, R.P., Gnanapragasam, P.N., and Bjorkman, P.J. (2012). Single-chain Fv-based anti-HIV proteins: potential and limitations. *J. Virol.* 86, 195–202.
- West, A.P., Jr., Scharf, L., Scheid, J.F., Klein, F., Bjorkman, P.J., and Nussenzweig, M.C. (2014). Structural insights on the role of antibodies in HIV-1 vaccine and therapy. *Cell* 156, 633–648.
- Witte, M.D., Theile, C.S., Wu, T., Guimaraes, C.P., Blom, A.E., and Ploegh, H.L. (2013). Production of unnaturally linked chimeric proteins using a combination of sortase-catalyzed transpeptidation and click chemistry. *Nat. Protoc.* 8, 1808–1819.
- Wolbank, S., Kunert, R., Stiegler, G., and Kattinger, H. (2003). Characterization of human class-switched polymeric (immunoglobulin M [IgM] and IgA) anti-human immunodeficiency virus type 1 antibodies 2F5 and 2G12. *J. Virol.* 77, 4095–4103.
- Wu, H., Pfarr, D.S., Tang, Y., An, L.L., Patel, N.K., Watkins, J.D., Huse, W.D., Kiener, P.A., and Young, J.F. (2005). Ultra-potent antibodies against respiratory syncytial virus: effects of binding kinetics and binding valence on viral neutralization. *J. Mol. Biol.* 350, 126–144.
- Wu, X., Yang, Z.Y., Li, Y., Hogerkor, C.M., Schief, W.R., Seaman, M.S., Zhou, T., Schmidt, S.D., Wu, L., Xu, L., et al. (2010). Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science* 329, 856–861.

- Yu, X., Yuan, X., Matsuda, Z., Lee, T.H., and Essex, M. (1992). The matrix protein of human immunodeficiency virus type 1 is required for incorporation of viral envelope protein into mature virions. *J. Virol.* 66, 4966–4971.
- Zadeh, J.N., Steenberg, C.D., Bois, J.S., Wolfe, B.R., Pierce, M.B., Khan, A.R., Dirks, R.M., and Pierce, N.A. (2011). NUPACK: Analysis and design of nucleic acid systems. *J. Comput. Chem.* 32, 170–173.
- Zhou, H.X. (2004). Polymer models of protein stability, folding, and interactions. *Biochemistry* 43, 2141–2154.
- Zhu, P., Liu, J., Bess, J., Jr., Chertova, E., Lifson, J.D., Grisé, H., Ofek, G.A., Taylor, K.A., and Roux, K.H. (2006). Distribution and three-dimensional structure of AIDS virus envelope spikes. *Nature* 441, 847–852.
- Zingler, K., and Littman, D.R. (1993). Truncation of the cytoplasmic domain of the simian immunodeficiency virus envelope glycoprotein increases env incorporation into particles and fusogenicity and infectivity. *J. Virol.* 67, 2824–2831.

Disease-Specific Alterations in the Enteric Virome in Inflammatory Bowel Disease

Graphical Abstract



Authors

Jason M. Norman, Scott A. Handley, ..., Miles Parkes, Herbert W. Virgin

Correspondence

virgin@wustl.edu

In Brief

The enteric virome is abnormal in multiple cohorts of inflammatory bowel disease patients, exhibiting disease-specific features that are not explained by changes in bacterial diversity and richness.

Highlights

- The enteric virome is abnormal in multiple inflammatory bowel disease patient cohorts
- The enteric virome richness increases in Crohn's disease and ulcerative colitis
- Decreases in bacterial diversity and richness in IBD do not explain virome changes
- Virome changes in Crohn's disease and ulcerative colitis are disease specific



Disease-Specific Alterations in the Enteric Virome in Inflammatory Bowel Disease

Jason M. Norman,^{1,10} Scott A. Handley,^{1,10} Megan T. Baldrige,¹ Lindsay Droit,¹ Catherine Y. Liu,¹ Brian C. Keller,^{1,2} Amal Kambal,¹ Cynthia L. Monaco,^{1,2} Guoyan Zhao,^{1,3} Phillip Fleshner,⁴ Thaddeus S. Stappenbeck,¹ Dermot P.B. McGovern,⁵ Ali Keshavarzian,⁶ Ece A. Mutlu,⁶ Jenny Sauk,⁷ Dirk Gevers,⁸ Ramnik J. Xavier,^{7,8} David Wang,^{1,3} Miles Parkes,⁹ and Herbert W. Virgin^{1,*}

¹Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO 63110, USA

²Department of Medicine, Washington University School of Medicine, St. Louis, MO 63110, USA

³Department of Molecular Microbiology, Washington University School of Medicine, Saint Louis, MO 63110, USA

⁴Division of Colorectal Surgery, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA

⁵The F. Widjaja Foundation Inflammatory Bowel and Immunobiology Research Institute; Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA

⁶Department of Medicine, Division of Digestive Diseases and Nutrition, Rush University Medical Center, Chicago, IL 60612, USA

⁷Gastrointestinal Unit and Center for the Study of Inflammatory Bowel Disease, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA

⁸Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

⁹Division of Gastroenterology Addenbrooke's Hospital and Department of Medicine, University of Cambridge, Cambridge CB2 0QQ, UK

¹⁰Co-first author

*Correspondence: virgin@wustl.edu

<http://dx.doi.org/10.1016/j.cell.2015.01.002>

SUMMARY

Decreases in the diversity of enteric bacterial populations are observed in patients with Crohn's disease (CD) and ulcerative colitis (UC). Less is known about the virome in these diseases. We show that the enteric virome is abnormal in CD and UC patients. In-depth analysis of preparations enriched for free virions in the intestine revealed that CD and UC were associated with a significant expansion of *Caudovirales* bacteriophages. The viromes of CD and UC patients were disease and cohort specific. Importantly, it did not appear that expansion and diversification of the enteric virome was secondary to changes in bacterial populations. These data support a model in which changes in the virome may contribute to intestinal inflammation and bacterial dysbiosis. We conclude that the virome is a candidate for contributing to, or being a biomarker for, human inflammatory bowel disease and speculate that the enteric virome may play a role in other diseases.

INTRODUCTION

Inflammatory bowel disease (IBD) is a complex, remitting and relapsing inflammatory disease with genetic and environmental risk factors. One environmental contributor is thought to be microorganisms that live in the intestine (Gevers et al., 2014; Kostic et al., 2014; Minot et al., 2011; Norman et al., 2014; Virgin, 2014). Of these microorganisms, bacteria have gained the greatest

attention and are linked to training mucosal immunity and minimizing mucosal inflammation (reviewed in Belkaid and Hand [2014]). An aberration in either of these immune processes can have detrimental consequences for IBD progression. For example, a reduction of *Bacteroidetes* and *Firmicutes* and expansion of normally less abundant bacterial taxa (dysbiosis), as well as changes in bacterial microbiome function, have been associated with both Crohn's disease (CD) and ulcerative colitis (UC) (Kostic et al., 2014; Stappenbeck et al., 2011). Importantly, household contacts without IBD can also exhibit signs of bacterial dysbiosis (Joossens et al., 2011). These individuals have increased intestinal permeability compared to healthy community controls (Hollander et al., 1986), suggesting that the bacterial microbiome is heavily influenced by the household environment. Investigations have also shown that the home environment is a primary determinant of the individual's bacterial microbiome and that humans are the primary vector of bacterial transmission between people living within the same household (Lax et al., 2014). Exchange of viruses between humans within a household has not been thoroughly investigated. Nevertheless, investigations of the bacterial microbiome and the enteric virome in IBD are likely to be optimized by the investigation of household controls rather than matched controls from different households.

Emerging data indicate that the viral component of the microbiome, termed the virome, can profoundly influence host physiology (Handley et al., 2012; Norman et al., 2014; Virgin, 2014). Recent advances in sequencing technology have led to the discovery of a diverse enteric human virome consisting of bacteriophages, as well as eukaryotic viruses (Breitbart et al., 2003; Finkbeiner et al., 2008; Minot et al., 2011, 2012, 2013; Reyes et al., 2010). Importantly, evidence that eukaryotic viruses can interact with IBD risk genes to alter intestinal disease comes from studies

of mice carrying mutations in *Il-10* or *Atg16l1*, indicating that members of the virome may contribute to IBD (Basic et al., 2014; Cadwell et al., 2010; Irving and Gibson, 2008; Sun et al., 2011). Bacteriophages may also play a direct role in intestinal physiology or change the bacterial microbiome through predator-prey relationships (Barr et al., 2013; Duerkop et al., 2012; Reyes et al., 2013; Willner et al., 2009, 2012).

The virome, much of which is composed of bacteriophages, contains the most diverse genetic elements on earth and is only beginning to be characterized at the sequence level (Virgin, 2014). In the absence of disease, enteric bacteriophage populations exhibit significant diversity between individuals and are temporally stable (Minot et al., 2013; Reyes et al., 2010). Bacteriophages in the healthy human intestine are predominantly temperate double-stranded DNA (dsDNA) *Caudovirales* or single-stranded DNA (ssDNA) *Microviridae* that latently infect their bacterial hosts and generate few viral progeny that may infect and kill other bacteria (Minot et al., 2011, 2013; Minot et al., 2011; Reyes et al., 2010; Waller et al., 2014). However, environmental stimuli, such as nitric oxide and antibiotics, induce the production of infectious bacteriophages that lyse their bacterial host and infect neighboring cells bearing specific receptors (Lindsay et al., 1998; Maiques et al., 2006; Zhang et al., 2000; Zhang and LeJeune, 2008). This process releases infectious virions into the intestine, which can be purified and analyzed. Alterations in bacteriophage abundance have been suggested in CD (Lepage et al., 2008; Pérez-Brocá et al., 2013; Wagner et al., 2013); however, these studies did not characterize the enteric virome in detail and did not control for factors within households that may influence the microbiome.

Here, we characterized the normal human and IBD enteric virome by metagenomic sequencing of the DNA of virus-like particle (VLP) preparations from fecal samples obtained from UC and CD patients and controls. Throughout the text, we refer to two ecological metrics, richness (the number of taxa counted per sample) and diversity. Diversity measures both richness and the relative abundance (or evenness) of the taxa present; changes in diversity can result from alterations in either richness or evenness. Detailed analysis of purified virions in VLP preparations and bacterial 16S ribosomal RNA sequences from a longitudinal patient cohort compared to household controls revealed the expected decrease in bacterial richness and diversity accompanied by a striking IBD-associated increase in bacteriophage richness. These findings were validated in two independent and geographically distinct patient cohorts that contained matched controls. The taxonomic substructure of the enteric virome and bacterial microbiome in CD and UC showed geographic variation in the specific bacteria and viruses detected. Together, these data support a model in which IBD-associated increases in bacteriophage richness are not merely accounted for by an increase in their bacterial host cells. We observed both positive and negative correlations between specific viral and bacterial taxa. These data demonstrate, for the first time, that unique changes in the bacteriophage component of the enteric virome occur in CD and UC, raising the possibility that these changes may contribute to disease pathogenesis, perhaps through a predator-prey relationship between bacteriophages and their bacterial hosts. These data provide a rationale

for considering virome diagnostics for IBD and manipulation of the enteric virome as a novel therapeutic strategy for the management of IBD and emphasize the need for a greater understanding of transkingdom interactions within the microbiome for other diseases associated with changes in the bacterial microbiome (Duerkop and Hooper, 2013; Norman et al., 2014; Virgin, 2014).

RESULTS

Virome Alterations Are Observed in Multiple Cohorts

To initially define the enteric virome associated with IBD, we performed metagenomic sequencing of stool filtrates using the Roche 454 platform on three independent cohorts consisting of IBD and non-IBD household controls (Tables S1 and S2; Cambridge, United Kingdom; Chicago, USA; and Los Angeles, USA). On average, we obtained $32,591 \pm 27,531$ sequences (number \pm SD) that were 282 ± 47 nucleotides in length from 72 fecal samples (Table S3; 12 household controls, 18 CD, and 42 UC). Sequences were demultiplexed, quality filtered, and assigned taxonomy (Supplemental Information). The majority of sequences obtained were assigned to the human host or to bacterial taxa (Table S3). Consistent with previous reports, bacteriophages of the *Caudovirales* order and *Microviridae* family were the most abundant viral taxa identified in all three cohorts (Figure 1A) (Minot et al., 2011; Reyes et al., 2010). Other viruses were detected in a limited number of samples and represented an average of five percent or fewer of the total viral sequences. An analysis of the relative abundances of sequences from all three cohorts revealed an inverse correlation between the *Caudovirales* and *Microviridae* (Figure 1B). This inverse correlation was also present when controls, CD, and UC were analyzed separately (data not shown). Comparing household controls to IBD samples within the UK cohort revealed that this disproportionate representation of bacteriophage abundance was associated with IBD (Figure 1C). Disparate ratios of *Caudovirales* and *Microviridae* were also observed in patients from Los Angeles. This suggestive correlation between disease and a change in sequences from the enteric virome was striking given the geographical and environmental diversity of the cohorts.

In-Depth Analysis of Free Virions in the Enteric Virome in IBD

These initial observations prompted us to perform an in-depth analysis of the virome by metagenomic sequencing of VLPs purified from the feces of patients and controls from 17 IBD households in the UK (Figure 2A and Tables S1 and S2). VLP purification enriches for free virions (Reyes et al., 2012; Thurber et al., 2009). To further refine the relationship between IBD and the enteric virome and to take into account prior data indicating that the bacterial microbiome and virome are similar within households (Lax et al., 2014; Reyes et al., 2010), we compared IBD patients to matched household controls (Figure 2A). This is particularly important for virome analysis given the high interpersonal variation of viruses (Reyes et al., 2010). Samples were collected from both the IBD patient and household control at the time of a clinical flare of disease (Supplemental Information). In total, 21 household control samples and 52 IBD samples



(B) Correlation plot of the *Caudovirales* and *Microviridae* relative abundance for all samples. Linear regression \pm 95% confidence interval and Spearman correlation coefficient are shown.

(C) *Microviridae* and *Caudovirales* relative abundance for United Kingdom household controls, UC and CD (top); Los Angeles and Chicago IBD (bottom). The bars indicate the median and interquartile range. Statistical significance was determined by the Mann-Whitney test.

See also [Tables S1, S2, and S3](#).

able to assign viral taxonomy to an average of $17,022 \pm 15,725$ sequences (15%) (Figure S1A), with the majority belonging to *Caudovirales* and *Microviridae* bacteriophages (Figure S1B). Sequences were also assigned to many less-abundant viral taxa, including bacteriophages whose annotated hosts include bacteria commonly found in human fecal samples and common eukaryotic viruses (Figure S1B). Analysis of our VLP sequences revealed a low level of contamination with human sequences

(24 active disease and 28 inactive disease samples; 36 total UC and 16 total CD) were used to isolate VLPs for sequencing. We validated observations in two additional cohorts from Chicago and Boston that contained CD and UC patients and matched healthy control subjects (described in greater detail below; [Tables S1](#) and [S2](#)).

For the UK cohort, we obtained $1,111,569 \pm 493,164$ paired-end sequences per sample with an average sequence quality of 36.5 ± 3.7 (Table S4 and Experimental Procedures). Quality control trimming resulted in an average of 2% reduction in the number of sequences to $1,094,360 \pm 503,337$ with an average reduction in sequence length from 250 to 241.7 ± 3.7 bases and an average increase in quality score of 1.0 ± 0.03 . There were no statistically significant differences in the number of total or quality-controlled sequences obtained between control and IBD patient samples (Figure S1A). Sequences were clustered at 95% identity to remove similar sequences and to generate sequences termed unique hereafter. Clustering resulted in an average reduction of 89% to $112,192 \pm 71,820$ unique sequences per sample. Interestingly, despite the fact that there were no significant differences in the total sequences obtained, we detected a significant increase in unique reads in CD patients compared to either household controls or UC patients. (Figure S1A).

Unique sequences were mapped to a custom virus protein database, as described in the [Experimental Procedures](#). We were

(0%– 4%). Possible contamination with bacterial sequences was confounded by the presence of integrated prophages in full genome sequences of bacteria. We mapped a subset of our total VLP enriched sequences from the UK cohort to the recently discovered crAssphage genome and detected it in 71% of our samples (Dutilh et al., 2014) (Table S4). The percentage of sequences that mapped to this virus varied greatly (range 1%–89%) and did not correlate with disease status or drug treatment.

Interestingly, we observed an increase in the richness of bacteriophages, specifically, members of the *Caudovirales* in IBD (Figures 3A and 3B). It is unlikely that these differences can be attributed to an uneven number of samples or unique sequences in control and disease groups because the rate of acquisition of new bacteriophage taxa in disease samples rapidly outpaces new taxa acquisition in control samples (Figure 3B). No increases in *Microviridae* richness or diversity were observed (Figure S2A), indicating that bacteriophage expansion was restricted to certain taxa and that our methods do not systematically increase bacteriophage richness in IBD samples compared to controls. Some IBD samples had many fewer *Microviridae* than controls, further supporting that the viromes were different between the two disease states (Figure S2A).

Taxonomic assignment of de novo assembled contigs longer than 1,000 nucleotides also indicated that *Caudovirales* were enriched in IBD (Figure 3C). These data indicated that the VLP preparations contained partial or complete virus genomes and that the overwhelming majority of those assignable genomes

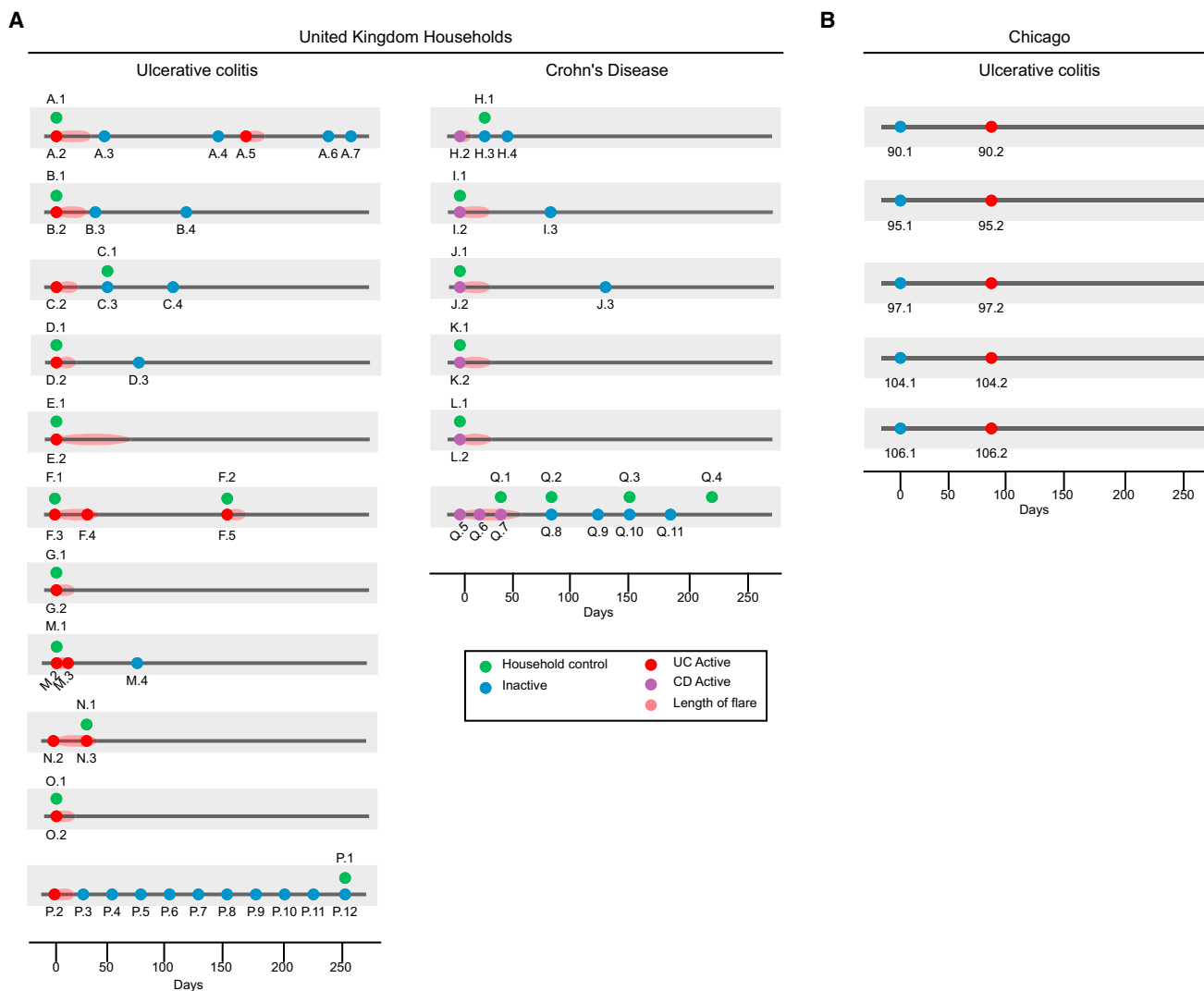


Figure 2. In-Depth, Longitudinal Cohort Graphical Timeline

(A) United Kingdom IBD stool samples and non-IBD household control stools were collected at the onset of symptoms (flare) and collected as symptoms resolved where indicated. The length of each IBD flare is indicated by the red shaded oval. The samples are annotated by household ID and sample number.

(B) Chicago UC samples were collected first during inactive disease and again when symptoms exacerbated. The samples are annotated by the subject and sample number.

See also [Tables S1](#) and [S2](#).

were from *Caudovirales* bacteriophages. Taken together, *Caudovirales* sequences and assembled contigs were differentially expanded in CD and UC compared to household controls and were therefore capable of having disparate effects on the microbiome and immune responses.

Disease-Specific Changes in the Enteric Virome

In addition to differences in bacteriophage richness between IBD and household control samples, we also observed striking differences in richness and the types of bacteriophage taxa observed between CD and UC samples ([Figure 3D](#)). A substantial number of taxa were observed among all samples; however, each disease type harbored unique bacteriophages. These

differences in *Caudovirales* taxa could be the result of multiple factors, including health or disease, household, duration of cohabitation, disease activity, age, sex, and/or drug treatment. We therefore performed a multivariate analysis of *Caudovirales* abundance using MaAsLin (Multivariate Association with Linear Models) on VLP sequences compared to their household controls ([Morgan et al., 2012](#)). MaAsLin identified 35 different *Caudovirales* that were significantly associated with different households (43 total associations) ([Tables S5](#) and [S6](#)). This relationship was also revealed by plotting *Caudovirales* relative abundances, which indicate conservation within households ([Figure 3E](#)). This finding is consistent with a previous report ([Reyes et al., 2010](#)). Five of the *Caudovirales* identified by

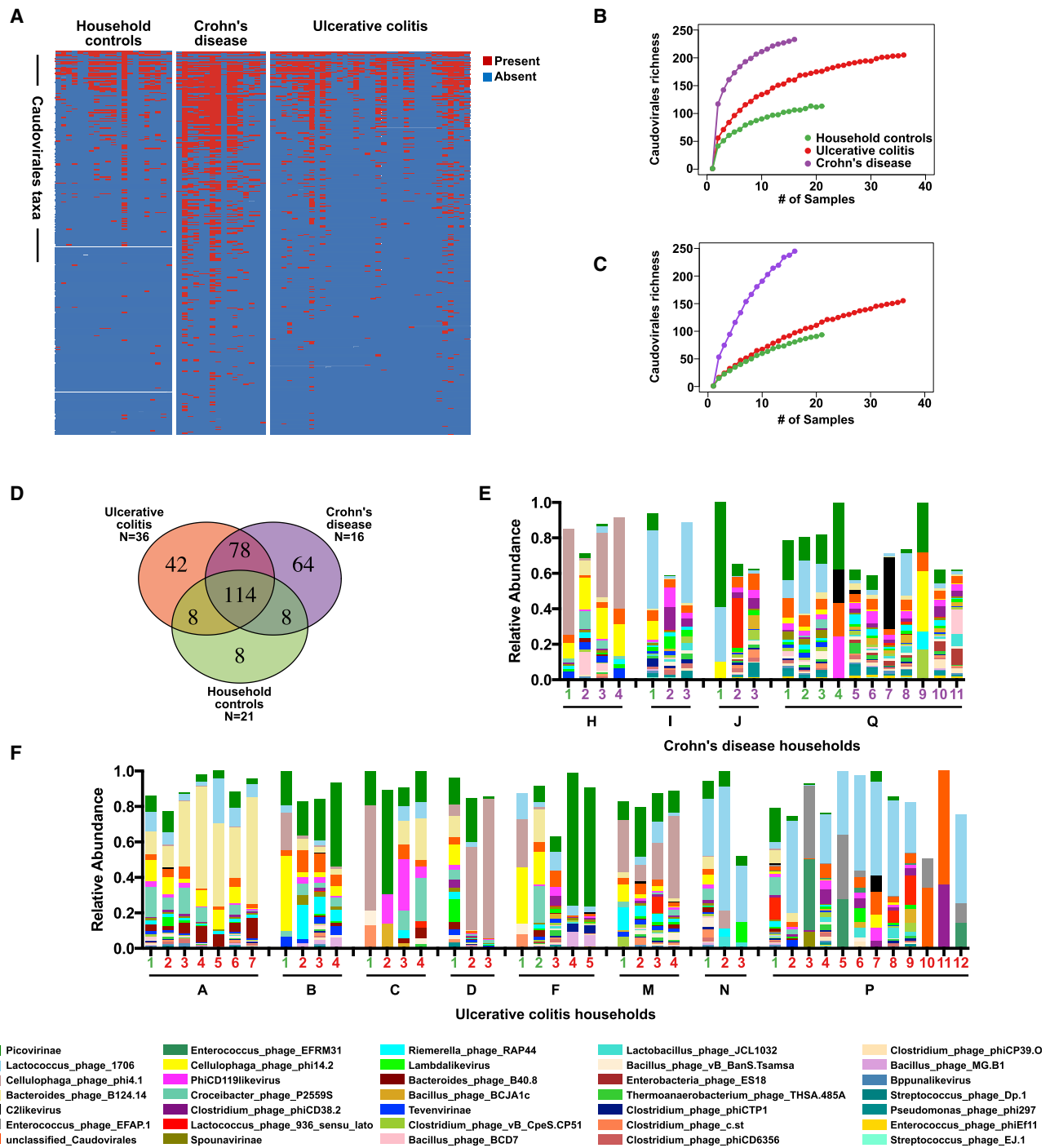
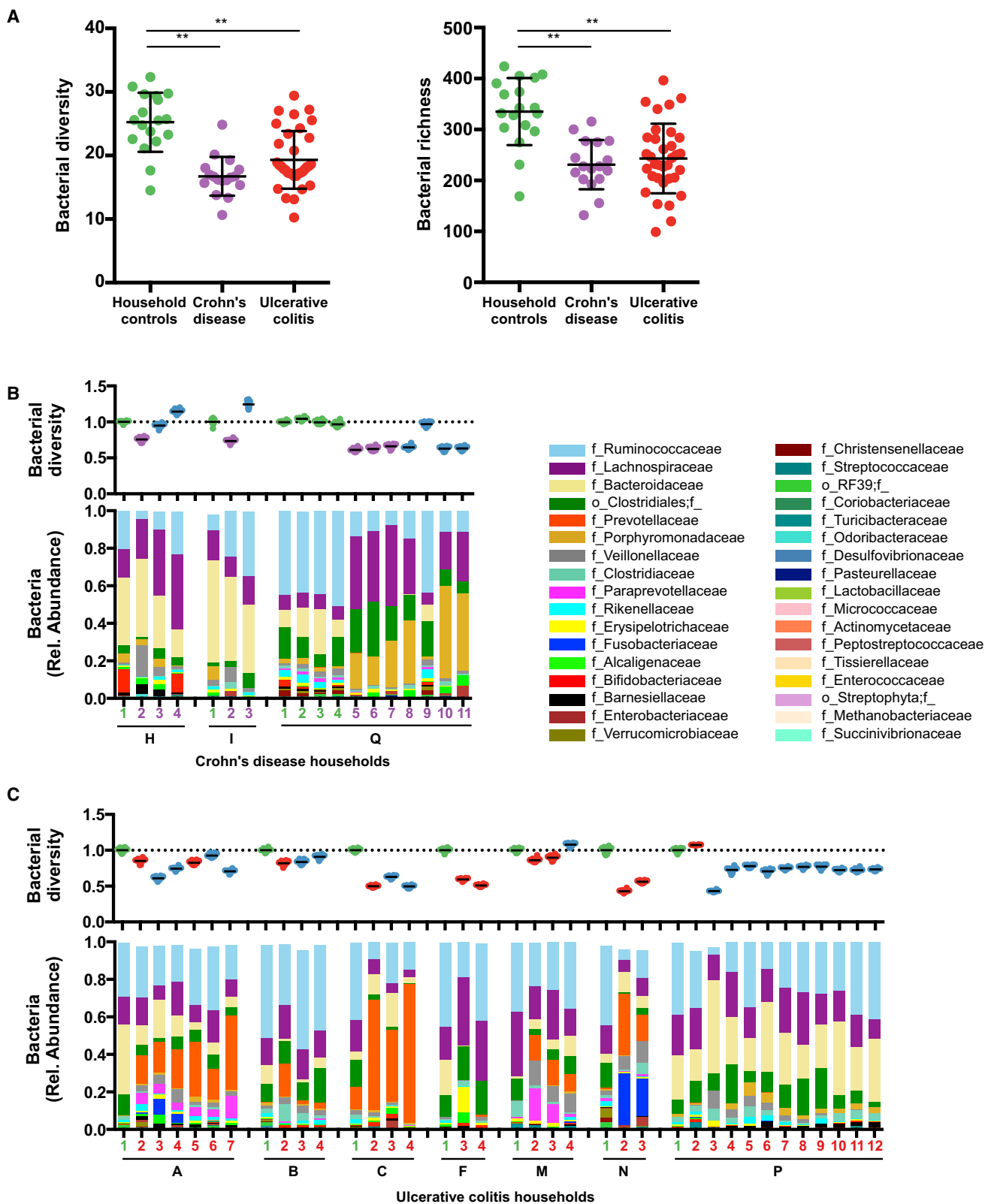


Figure 3. Bacteriophage Expansion Is Associated with CD and UC

(A) Presence-absence heat map of the sequences assigned to *Caudovirales* taxa in VLP preparations from UK household control, CD and UC stool samples. (B and C) Rarefaction curves of *Caudovirales* richness versus an increasing number of subsamplings with replacement. (B) *Caudovirales* richness based on individual sequences. (C) Richness based on assembled *Caudovirales* contigs. The curves represent the average of 500 iterations at each depth of samples. (D) Venn diagram of the *Caudovirales* taxa in household control, CD and UC samples. n indicates the number of samples within each subgroup. (E and F) Plots of the relative abundance of the 35 most abundant *Caudovirales* taxa in the UK UC and CD households. Bars are annotated by the household ID and sample number. Green numbers indicate household controls; purple numbers indicate CD; red numbers indicate UC. See also Figure S1 and Tables S4, S5, and S6.



(legend on next page)

MaAsLin were significantly associated with disease, including sequences most closely related to *Lactococcus*, *Lactobacillus*, *Clostridium*, *Enterococcus*, and *Streptococcus* bacteriophages (Table S6). The increased *Caudovirales* richness in CD samples relative to household controls corresponded with increased bacteriophage diversity (Figure S2B). This was not observed in UC, further highlighting differences in the virome between UC and CD.

Inverse Correlation between IBD-Associated Changes in the Virome and Bacterial Microbiome

We next assessed whether the increase in bacteriophage richness observed in CD and UC was associated with a parallel change in bacterial populations. To answer this question, we performed bacterial 16S rRNA gene sequencing (Tables S1 and S2). On average, we obtained $64,107 \pm 38,570$ sequences that were clustered at 97% identity into $56,096 \pm 32,774$ operational taxonomic units (OTUs) per sample (Table S7). As expected, CD and UC were associated with a significant reduction in bacterial diversity and bacterial richness compared to household controls (Figures 4A and 4B). However, we also observed significant similarity in the bacterial microbiome within IBD households as determined by permutation multivariate analysis of variance of the weighted UniFrac distances (Figures 4B and 4C; ADONIS $p = 0.001$, 999 permutations) (McArdle and Anderson, 2001). Interestingly, in patients that were sampled longitudinally, the bacterial diversity did not recover during periods of disease inactivity in either CD or UC (Figures 4B and 4C).

Like the shifts observed for bacteriophages, the changes in bacterial diversity could be multifactorial. We therefore performed a multivariate analysis of bacterial abundance using MaAsLin (Morgan et al., 2012). We identified 18 different bacterial taxa that were significantly associated with disease or disease activity (Tables S5 and S8). The vast majority of the significant OTUs were of the *Bacteroidetes* and *Firmicutes* phyla, including significant differences in members of the bacterial families *Ruminococcaceae*, *Lachnospiraceae*, *Bacteroidaceae*, and *Prevotellaceae*. Therefore, and as expected, disease-specific alterations in bacterial taxa were observed in our cohort; however, the variation was also largely linked to specific households, supporting previous studies of the microbiome that have indicated similarities among IBD patients and their household contacts (Joossens et al., 2011).

Importantly, the observed reduction in bacterial diversity was inversely related to the expansion of *Caudovirales* bacteriophages in IBD. Comparing the bacterial and *Caudovirales* bacteriophage communities in household controls, CD and UC samples indicated clear differences between the disease states (Figure 5A). Among the UK cohort, significant positive correla-

tions were observed in five out of six possible comparisons between bacterial diversity or richness and *Caudovirales* richness or diversity. However, a significant inverse correlation was observed in CD samples between *Caudovirales* diversity and both bacterial richness and diversity, which suggests that bacteriophage expansion was not simply the result of increases in their bacterial hosts (Figure 5A).

To further characterize the relationships between *Caudovirales* and bacterial taxa, we calculated the Spearman correlation between the *Caudovirales* taxa and the bacterial families found to be significantly altered in disease by MaAsLin analysis. Similar to the overall diversity and richness correlations, inverse correlations between *Caudovirales* and the significantly altered bacterial taxa are prevalent in UK CD patients (Figure 5B). In particular, the *Bacteroidaceae* bacterial families were inversely correlated with several *Caudovirales* taxa in CD. This corresponded with a reduction in the relative abundance of these bacterial taxa in CD patients compared to household controls. In contrast, the *Caudovirales* were positively correlated with the *Enterobacteriaceae* and *Pasteurellaceae* bacterial families in CD; these bacterial families were increased in abundance in CD patients (Figure 5B). Positive correlations were also observed between the *Caudovirales* and *Prevotellaceae* in CD; however, there were no changes in the relative abundance of *Prevotellaceae* in CD, further indicating that we were not merely sequencing prophages. Fewer positive or negative correlations were observed in UC patients despite the significant expansion of *Caudovirales* bacteriophages and decreased bacterial diversity in these patients (Figures 3B and 4A), further suggesting the existence of disease-specific elements of the virome-bacterial microbiome relationship in UC versus CD.

Independent IBD Cohorts for Validation of Virome Findings

We considered the possibility that our observations in UK IBD patients were either geographically determined or not reproducible. Therefore, we performed bacterial 16S and VLP sequencing of stool samples from two additional cohorts of CD and UC patients (Figure 2 and Tables S1 and S2). In these cases, household controls were not available, and so matched healthy controls were used. The first validation cohort was from Chicago and included 23 healthy control samples and 25 IBD samples (18 UC and 7 CD), including several samples that were part of the initial 454 analyses (Figure 1). For five of the Chicago UC patients, we were able to acquire samples both during inactive and active disease (Figure 2B). Two household controls that matched UC patients from Chicago were included in these analyses as healthy subjects. The second validation cohort was from Boston and included 10 healthy control samples and 25 IBD samples (11 UC and 14 CD).

Figure 4. Alterations in the Bacterial Community Composition in IBD

(A) Alpha diversity (left) and bacterial species richness (right) based on 16S rRNA gene sequences in the stool of household controls, CD, and UC patients. Error bars indicate the mean \pm SD. Statistical significance was determined by the Kruskal-Wallis test with Dunn's correction comparing all samples to all samples. ** $p > 0.01$.

(B and C) Plots of alpha diversity normalized to the diversity in household controls (top) and relative bacterial family abundance (bottom) of UK (B) CD households and (C) UC households. Green numbers indicate household controls; purple numbers indicate CD; red numbers indicate UC.

See also Tables S5, S7, and S8.

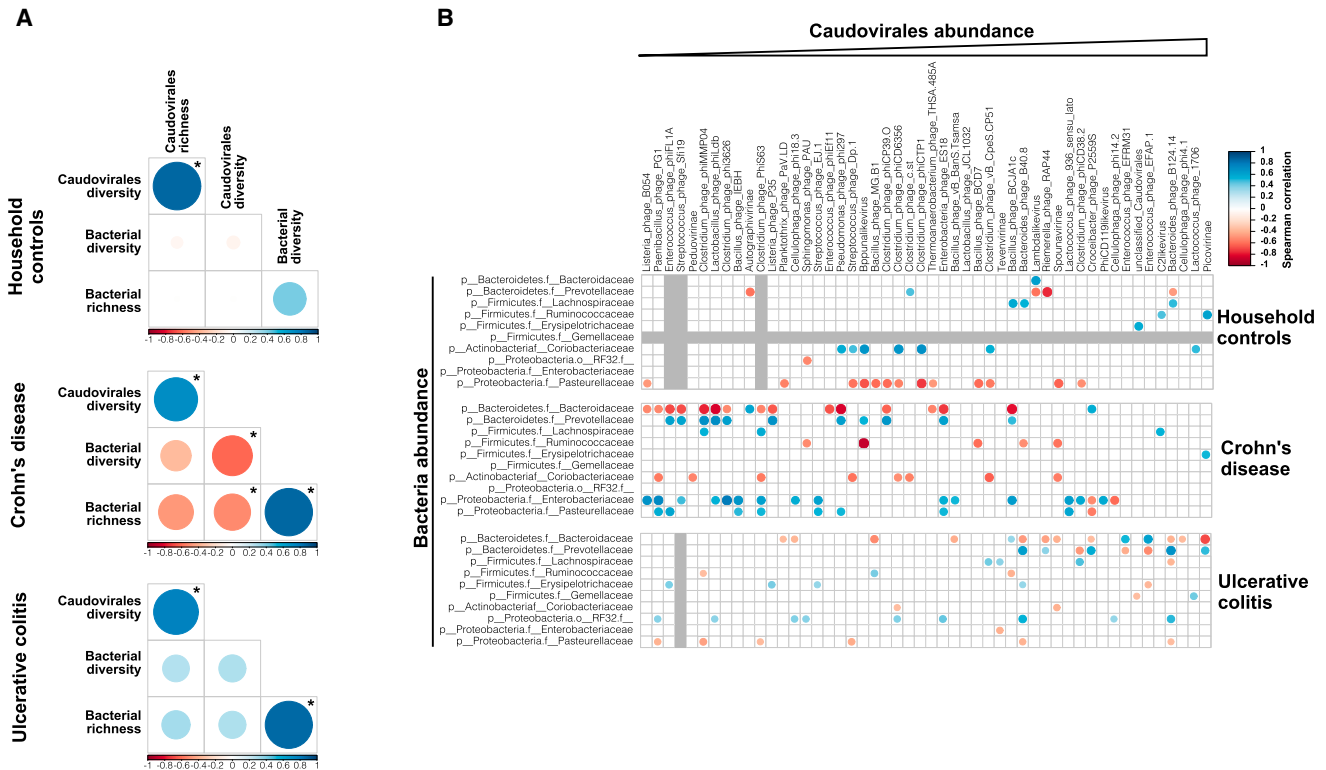


Figure 5. Disease-Specific Bacteria-Caudovirales Patterns in IBD

(A) Spearman correlation plot of *Caudovirales* richness, *Caudovirales* Shannon diversity, bacterial alpha diversity, and bacterial species richness for UK household control, CD, and UC samples. Statistical significance was determined for all pairwise comparisons; those with p values < 0.05 are indicated. Positive values (blue circles) indicate positive correlations, and negative values (red circles) indicate inverse correlations. The size and shading of the circle indicate the magnitude of the correlation where darker shades are more correlated than lighter shades.

(B) Spearman correlation plots of the relative abundances of the 50 most abundant *Caudovirales* taxa and bacterial families identified to be significantly associated with disease. UK household control (top), CD (middle), and UC (bottom) samples. The gray bars indicate any taxa that were not detected in the cohort subgroup. Statistical significance was determined for all pairwise comparisons; only significant correlations (p value < 0.05) are displayed.

See also Figure S2.

Relationship between *Caudovirales* Richness and IBD across Cohorts

As expected, significant reductions in bacterial diversity and richness were observed in CD and UC patients compared to healthy controls from both the Chicago and Boston validation cohorts (Figures 6A–6C). MaAsLin analysis of bacterial taxa revealed significant associations with IBD and disease activity in both Chicago and Boston cohorts, although many fewer bacterial taxa were significantly associated with IBD than observed in the UK cohort, which included household controls (Tables S5 and S8).

As observed in the UK cohort (Figure 3), a significant expansion of *Caudovirales* bacteriophages was observed in IBD patients from both validation cohorts (Figure 7). However, the specific relationships between bacteriophage richness and disease varied between cohorts. In Chicago, CD patients were *Caudovirales* rich compared to healthy controls (Figures 7A and 7B). This was evident when richness was assessed for both individual sequences and *Caudovirales* contigs. In the Boston cohort, both CD and UC patients had increased *Caudovirales*, with UC contigs being more enriched than CD (Figures 7A and 7B). We also detected unique *Caudovirales* taxa in CD and UC samples

from both the Chicago and Boston cohorts (Figure 7C), which was anticipated given our observations in the UK samples and the high inter-individual virome diversity reported previously (Reyes et al., 2010). Multivariate analysis of the relative abundance of *Caudovirales* in the Chicago and Boston samples revealed several disease-specific associations (Tables S5 and S6). We were unable to complete a correlation analysis to associate bacteriophage to bacteria abundances (as in Figure 5) due to the inadequate number of bacterial taxa associated with disease diagnosis as determined by MaAsLin analysis (Table S8). Therefore, we were unable to validate specific relationships between bacteriophage and bacterial taxa across our cohorts. Together, these data indicate that *Caudovirales* bacteriophages were expanded in both CD and UC patients compared to household or healthy controls from three independent cohorts.

Eukaryotic Viruses in IBD Cohorts

We also took advantage of the available data from the three IBD cohorts to analyze sequences from eukaryotic viruses. Anellovirus sequences were more prevalent in IBD samples compared to healthy controls (Table S4; anellovirus positive: household

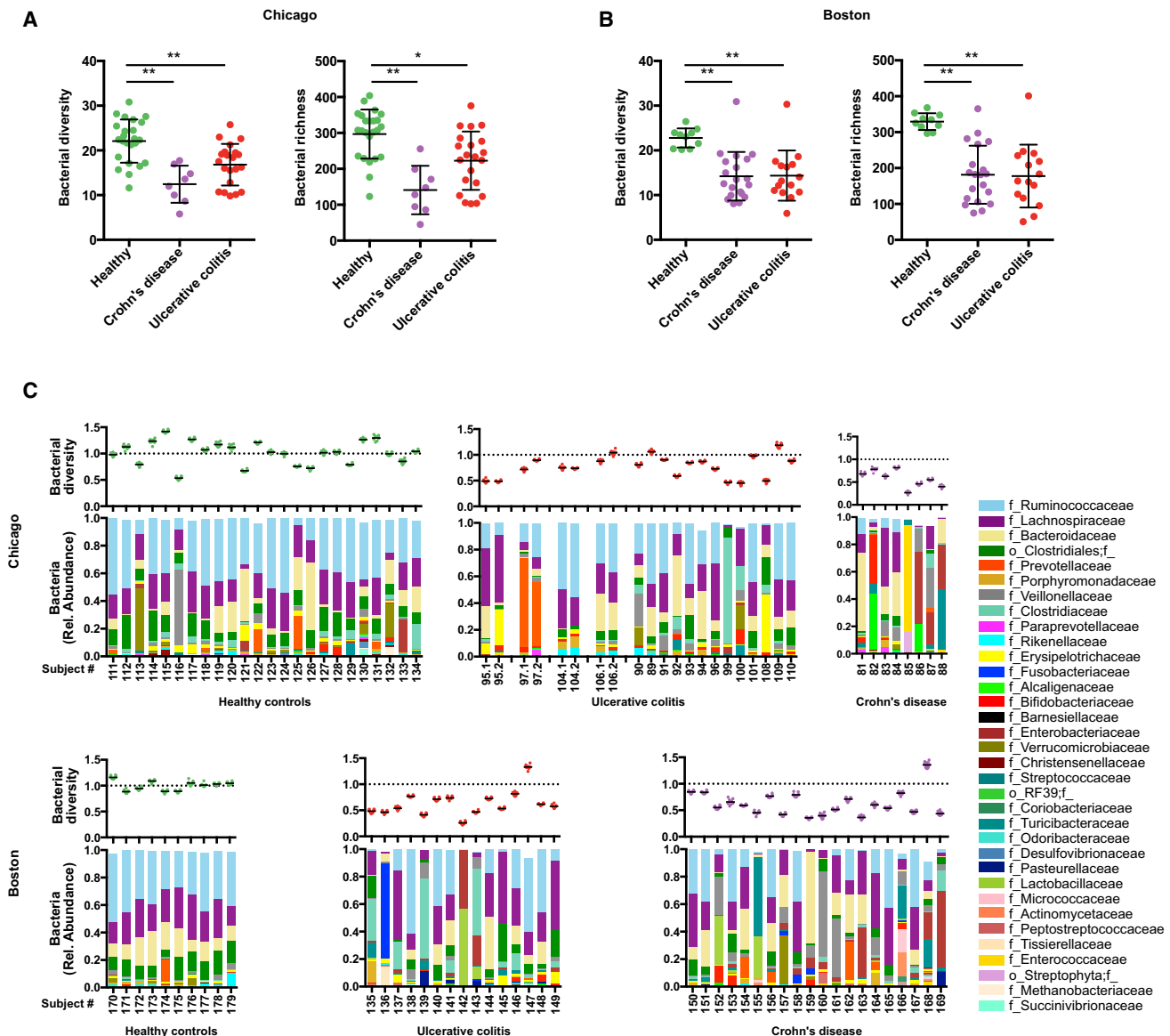


Figure 6. Validation of Bacterial Dysbiosis in Two Additional Cohorts

(A and B) Faith's phylogenetic alpha diversity (left) and bacterial species richness (right) based on 16S rRNA gene sequences in the stool of (A) Chicago healthy controls, CD, and UC patients and (B) Boston healthy controls, CD, and UC patients. Error bars indicate the mean \pm SD. Statistical significance was determined by the Kruskal-Wallis test with Dunn's correction comparing all samples to all samples. * $p > 0.05$ and ** $p > 0.01$.

(C) Plot of alpha diversity normalized to the average diversity in healthy subjects (top) and relative bacterial family abundance (bottom) for the Chicago and Boston cohorts. Healthy control, UC, and CD samples are indicated. Longitudinal samples from the same subject are grouped together.

See also [Tables S5](#), [S7](#), and [S8](#).

controls, 0%; healthy controls, 4.7%; CD, 27%; UC, 29%). However, anellovirus sequences from fecal samples were not detected in all patients and did not correlate with disease activity or drug treatment.

DISCUSSION

In this paper, we demonstrate that disease-specific changes in the enteric virome occur in both major forms of IBD, Crohn's dis-

ease, and ulcerative colitis. This was observed in a cohort of patients in comparison to household controls with increased power to detect disease-associated changes in the metagenome and validated in two independent and geographically distinct cohorts containing matched controls. The primary change in the virome associated with IBD was a significant expansion of the taxonomic richness of *Caudovirales* bacteriophages. Importantly, although this change was observed in both CD and UC, the viruses responsible for the change appeared to differ between

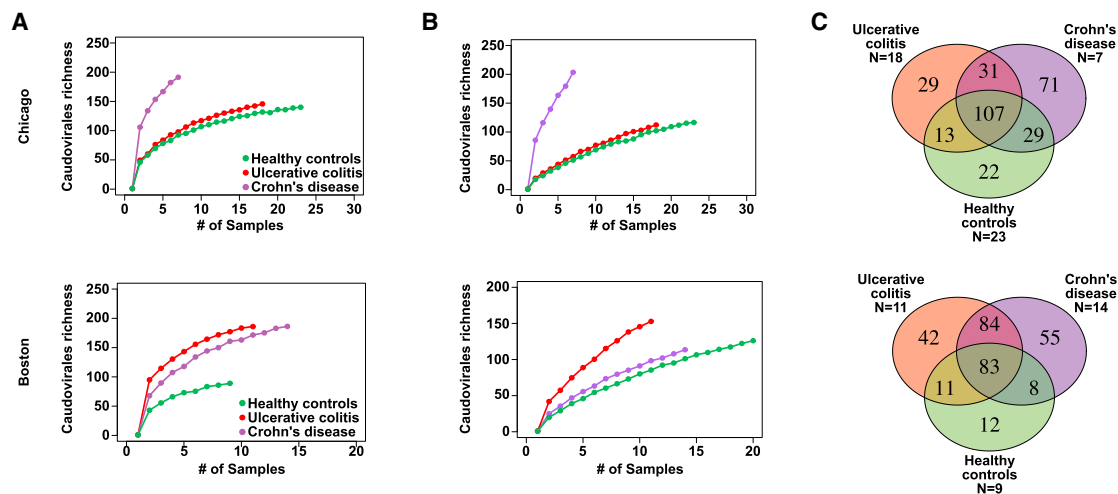


Figure 7. Validation of *Caudovirales* Expansion in Two Additional Clinical Cohorts

(A and B) Rarefaction curves of *Caudovirales* richness versus an increasing number of subsamplings with replacement for the Chicago (top) and Boston (bottom) cohorts.

(A) *Caudovirales* richness based on individual sequences.

(B) Richness based on *Caudovirales* contigs.

(C) Venn diagram of the *Caudovirales* taxa in healthy control, CD, and UC samples from Chicago (top) and Boston (bottom). n indicates the number of samples within each subgroup.

See also Tables S4, S5, and S6.

the two diseases, suggesting that the virome is specific for CD versus UC. Comparison across cohorts revealed that enteric viromes were unique between individuals and between cohorts, which is consistent with previous reports on the incredible diversity of the human gut virome (Minot et al., 2011; Reyes et al., 2010). Although the variability of gut bacteria and viruses in non-household controls and IBD patients makes it more difficult to observe specific relationships between individual bacteriophages and individual bacterial taxa, an increase in bacteriophage richness was consistently associated with IBD despite a decrease in bacterial richness and diversity.

Our data are consistent with reports that detected more *Caudovirales* bacteriophage sequences in intestinal washes and biopsy tissues of pediatric CD patients compared to non-inflammatory controls (Wagner et al., 2013) and enumerated more *Caudovirales* virions in CD biopsy washes by microscopy (Lepage et al., 2008). We believe that our observations reflect an expansion of infectious bacteriophages in IBD for two reasons. First, we sequenced VLP preparations enriched for virions. Second, an expansion of temperate bacteriophages integrated in bacterial genomes would be predicted to positively correlate with their bacterial hosts while we observe an inverse correlation. However, our data do not rule out the possibility that bacterial species harboring specific prophages are also expanded. Analysis of this will require full sequencing of the bacterial microbiome to complement analysis of bacterial taxa via analysis of 16S ribosomal RNA sequences.

Potential Role of Bacteriophages in IBD

In the human gut and in many ecosystems, the predominant bacteriophages are tailed, dsDNA *Caudovirales* and non-tailed,

ssDNA *Microviridae* (Breitbart et al., 2003; Reyes et al., 2010). The biology of bacteriophages has been extensively reviewed (Brüssow et al., 2004; Clokie et al., 2011; Fortier and Sekulovic, 2013). The expansion in *Caudovirales* bacteriophage richness observed here could arise from the induction of prophage from commensal microbes or reflect the introduction of new viruses acquired from the environment, for example from food or contact with other people, including household contacts. These changes might have significant consequences for the bacterial microbiome. For example, bacteriophages are primary drivers of bacterial diversity and fitness in different ecosystems (Brüssow et al., 2004). In the gut, bacteriophages are responsible for the horizontal transfer of genetic material among bacterial communities, including those for pathogenesis (e.g., cholera toxin, pertussis toxin, and shiga toxin) and antibiotic resistance (Brüssow et al., 2004; Maiques et al., 2006; Zhang and LeJeune, 2008). Widespread bacteriophage induction, mutation, or introduction from external sources could effectively shuffle the deck of bacterial fitness and resistance genes. Second, the activation of latent prophages leads to the lysis of their bacterial hosts and can have profound ecological consequences (Weitz and Wilhelm, 2012). The intestinal microbiome has been shown to be sensitive to bacteriophage invasion, leading to changes in the abundance of specific gut bacterial species (Reyes et al., 2013). Lysis of bacteria would also be expected to release proteins, lipids, and nucleic acids that serve as pathogen-associated molecular patterns (PAMPs) and antigens that trigger inflammatory signaling cascades to induce cytokines, cellular infiltration, and tissue damage. The development of animal models to test these predator-prey relationship(s) in IBD pathogenesis and intestinal inflammation will certainly be an important area for future investigation.

Another potential consequence of a change in enteric bacteriophages might result from direct interactions between these viruses and the mammalian host. For example, bacteriophages are able to translocate from the gastrointestinal (GI) lumen to systemic sites in animals (Górski et al., 2006), CD patients, and healthy controls (Parent and Wilson, 1971). Bacteriophages are also capable of inducing humoral immune responses (Uhr et al., 1962). Further, in vitro stimulation of macrophages with bacteriophage particles induces MyD88-dependent proinflammatory cytokine production (Eriksson et al., 2009). Chronic intestinal inflammation is the most basic element of IBD pathology, leading to the destruction of intestinal tissue and increased epithelial permeability. This leads to increased systemic exposure to the flow of microbial immunogens, potentially including those from bacteriophages and lysed bacterial cells, to maintain and further exacerbate inflammation. For these reasons, bacteriophages may serve as antigens or innate immune ligands that stimulate host immunity and inflammation.

The “Dark Matter” of the Virome

The metagenomic sequencing of samples enriched for intact viroids has led to an appreciation of the incredible richness of enteric viruses in humans. Importantly, the methods we used here would not be expected to detect either enveloped viruses or RNA viruses and so there may be much to learn about these aspects of the IBD virome. Prior attempts to characterize the virome using these methods were only able to assign 60% to 87% of VLP sequences or contigs to anything within sequence databases (Minot et al., 2011, 2013). Across our cohorts, we were only able to assign on average 14% of VLP sequences to a viral database. The apparent discrepancy between our study and previous ones may be explained by our use of higher throughput Illumina-based sequencing, sequence databases, or taxonomic assignment criteria. The percent identity of those VLP sequences that were assigned in this study varied greatly in a subset of sequences that were analyzed (40% to 100% identity; data not shown). This is a major issue for future studies, as we were only able to report on the bacteriophages that are most closely related to taxa in the database. It is likely that additional viruses are present in our sequencing data sets but are not detected due to this limitation. An implication of this is that current databases lack sufficient depth for us to be able to link specific bacteriophages to individual bacteria or disease. This limitation will only be overcome by significant expansion of the bacteriophage sequences and annotations in available databases. This will not be a simple task and will require a global, coordinated effort to improve virus databases; it is clear that the Human Microbiome Project (<http://www.hmpdacc.org/>), an effort of similar scale, has improved the ability to classify and study bacteria. As databases improve, the resolution of our picture of bacterial microbiome-virome associations will improve. Many of the sequences that we did not assign using our viral database map to bacterial genomes (data not shown), an assignment complicated by the fact that many bacterial genome sequences contain prophages that have not been independently annotated. Addressing the misannotation of integrated prophages as “bacterial” sequences will require development of new tools and significant improvement of viral databases containing

a much greater diversity of fully annotated complete viral genomes.

Virome Implications for Disease Pathogenesis, Treatment, and Monitoring

The primary therapeutic goal in IBD is to limit inflammation and halt or even reverse tissue destruction. In many cases, pharmacologic or biomolecule therapies fail, and surgical resection of inflamed/damaged tissue is required. Thus, novel approaches are required to optimize IBD management. One potential approach is the manipulation of enteric bacteria through probiotics and prebiotics, which have had limited success in humans thus far (Butterworth et al., 2008; Naidoo et al., 2011; Shanahan and Quigley, 2014). Fecal transplantation from healthy donors or with defined bacterial cultures is an approach that is gaining traction due to its success in the treatment of recurring *Clostridium difficile* infections (van Nood et al., 2013). However, early attempts at curing bacterial dysbiosis in UC through fecal transplantation have not proven successful (Angelberger et al., 2013; Kump et al., 2013), perhaps due to the instability of the donor microbiome in IBD patients. It is intriguing to speculate that a disease-associated, taxonomically rich virome in the recipient may interact with donor microbes to limit probiotic or fecal transplantation efficacy. Defining the virome before and during probiotic or fecal transplantation will be required to assess this possibility. It is notable that, in animal models, eukaryotic viruses change intestinal biology and inflammation by acting in concert with host bacteria in a manner dependent on host genetics (Basic et al., 2014; Cadwell et al., 2010). Thus both eukaryotic viruses and bacteriophages may have a role in IBD through interactions with the bacterial microbiome. It will be important to more completely understand the interactions between bacteria and viruses and the host to be able to develop and personalize these approaches to managing IBD.

Our data also suggest that the specific expansion of bacteriophages in CD is associated with decreased bacterial diversity. The development of methods that block the infection of their bacterial hosts by these bacteriophages is worth investigation. Furthermore, the identification of disease-specific *Caudovirales* could prove useful in differentiating CD and UC in the ~10% of cases of IBD in which the clinical phenotype is indeterminate between the two. Here, we did not observe any significant changes in the virome in IBD patients over time as disease activity changed. However, larger cohorts with more frequent repeated sampling of both IBD patients and household controls are required to more fully assess enteric virome stability in IBD.

Although IBD is associated with bacterial dysbiosis, additional—and much more common—diseases are also associated with changes in the bacterial microbiome. These include diabetes, obesity, metabolic diseases, and cancer (Larsen et al., 2010; Ley et al., 2005; Nicholson et al., 2012; Qin et al., 2012; Shen et al., 2010; Turnbaugh et al., 2006, 2009; Zackular et al., 2013). Our data indicate that understanding the bacterial microbiome in these diseases likely requires concurrent analysis of the virome. More speculatively, we question whether bacterial microbiome changes in many diseases are secondary to changes in emergence of temperate bacteriophages or introduction of bacteriophages from food, the environment, or through human

or animal contact. Thus, data presented here identifying inverse relationships between the bacterial microbiome and the enteric virome open up a new area of research in IBD and perhaps other diseases that have been shown to be associated with changes in the bacterial microbiome.

EXPERIMENTAL PROCEDURES

Cohort Description

Stool samples were collected at four gastroenterology clinics: (1) Addenbrooke's Hospital, University of Cambridge, UK; (2) Cedars-Sinai Hospital, Los Angeles, USA; (3) Rush University Medical Center, Chicago, USA; and (4) Massachusetts General Hospital, Boston, USA. A detailed list of the subjects and a full description of each cohort are included in the [Supplemental Information](#).

Virus-like Particle Enrichment and Sequencing

Virus-like Particle Preparation

VLPs were enriched from pulverized human stool using a protocol based on previously described methods (Reyes et al., 2010; Thurber et al., 2009). Stool sample filtrates were treated with lysozyme and chloroform to degrade any unfiltered bacterial and host cell membranes. Nonvirus protected DNA was degraded by treating with a DNase cocktail followed by heat inactivation of DNases. VLPs were lysed and nucleic acid was extracted.

VLP DNA was amplified, and DNA was randomly fragmented by ultrasonication before Illumina library construction. An equimolar pool of 12 samples was sequenced on an Illumina MiSeq instrument.

Sequence Processing and Analysis

Adapters and low-quality bases were trimmed, and reads were clustered at 95% identity. Unique sequences were queried against a customized viral database using BLASTx. Reads were assigned taxonomy using the lowest-common ancestor algorithm as implemented in MEGAN (v5.1.5) (Huson et al., 2011). Absolute read counts for selected viral taxa were exported from MEGAN and imported into R, data were normalized, and richness and diversity were calculated.

De Novo Contig Assembly

Contigs were assembled using the IDBA_UD assembler (v 1.1.0) using minimum and maximum kmer lengths of 20 and 120, respectively (Peng et al., 2012). All contigs larger than 1,000 nucleotides were compared to a viral genome reference sequence database consisting of 5,500 viral genomes available in NCBI as of July 7, 2014, using dc_megablast.

Bacterial 16S rRNA Analysis

16S rRNA Gene Analysis

Stool total nucleic acid was extracted from aliquots of pulverized human stool, as previously described (Reyes et al., 2013) with minor modifications for lower throughput processing of human stool ([Supplemental Information](#)). Primer selection and PCR was performed following previously described methods (Caporaso et al., 2011). The final pooled samples were sequenced on the Illumina MiSeq platform. 16S analysis was done in QIIME (Quantitative Insights into Microbial Ecology, version 1.8.0) (Caporaso et al., 2010), and OTU relative abundance, diversity and richness plots were generated in GraphPad Prism (version 6.0d).

ACCESSION NUMBERS

The MG-RAST project ID for the Roche 454 sequences reported in [Figure 1](#) of this paper is 11446. The EMBL-EBI accession number for the VLP and 16S sequences reported in this paper is PRJEB7772.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, two figures, and eight tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.01.002>.

AUTHOR CONTRIBUTIONS

J.M.N., S.A.H., H.W.V., M.P., A. Keshavarzian, E.A.M., D.P.B.M., and T.S.S. contributed to study conception and design. Recruitment of study subjects and acquisition of samples was performed by M.P., D.P.B.M., P.F., A.K., E.A.M., L.D., D.W., J.S., D.G., and R.J.X. J.M.N., S.A.H., M.T.B., B.C.K., L.D., C.L.M., and C.Y.L. contributed to acquisition of data. Analysis and interpretation of data was conducted by J.M.N., S.A.H., G.Z., and H.W.V. J.M.N., S.A.H., A. Kambal, and H.W.V. drafted the manuscript. J.M.N., S.A.H., H.W.V., M.P., D.P.B.M., A. Keshavarzian, E.A.M., T.S.S., D.W., A. Kambal, D.G., and R.J.X. contributed to critical revision.

ACKNOWLEDGMENTS

H.W.V., S.A.H., and J.M.N. were supported by R01 AI084887, R01 OD011170, grants 3132 and 274415 from the Crohn's and Colitis Foundation, and grant IBD-0357 from the Broad Medical Foundation. T.S.S. was supported by grant 3132 from the Crohn's and Colitis Foundation. J.M.N. was supported by NIH grant 5T32AI007163-35. M.T.B. was supported by NIH grant 5T32CA009547. B.C.K. was supported by NIH grant T32 HL007317-36. C.L.M. was supported by NIH grant 5T32AI007172-34. A. Keshavarzian and E.A.M. were supported by R01 AT007143, DK071838, and AT001628. D.P.B.M. was supported by DK062413, DK046763-19, AI067068, and U54DE023789-01, grant HS021747 from the Agency for Healthcare Research and Quality, grant 305479 from the European Union, and The Leona M. and Harry B. Helmsley Charitable Trust. M.P. was supported by the NIHR Biomedical Research Centre awards to Addenbrooke's Hospital and University of Cambridge School of Clinical Medicine. D.W. was supported by the NIH Midwest Regional Center of Excellence for Biodefense and Emerging Infectious Diseases Research grant U54 AI057160. D.G. and R.J.X. were supported by U54 DE023798 and R01 DK092405. We thank the Genome Technology Access Center in the Department of Genetics at Washington University School of Medicine for help with genomic analysis. The Center is partially supported by NCI Cancer Center Support grant P30 CA91842 to the Siteman Cancer Center and by ICTS/CTSA grant UL1TR000448 from the National Center for Research Resources (NCRR), a component of the NIH, and NIH Roadmap for Medical Research. This publication is solely the responsibility of the authors and does not necessarily represent the official view of NCRR or NIH. We would like to thank the patients and relatives who contributed samples to this collection and the team of research nurses and coordinators who supported this activity, including Claire Dawson, Elizabeth Andersen, and Martina Lofthouse (Cambridge, UK) and Gildardo Barron (Los Angeles, USA). We also acknowledge the NIHR Biomedical Research Centre award to support research infrastructure at Addenbrooke's Hospital/University of Cambridge School of Clinical Medicine. We would also like to thank Jessica Hoisington-Lopez from the Center for Genome Sciences and Systems Biology at Washington University School of Medicine for her sequencing expertise.

Received: May 17, 2014

Revised: October 13, 2014

Accepted: December 24, 2014

Published: January 22, 2015

REFERENCES

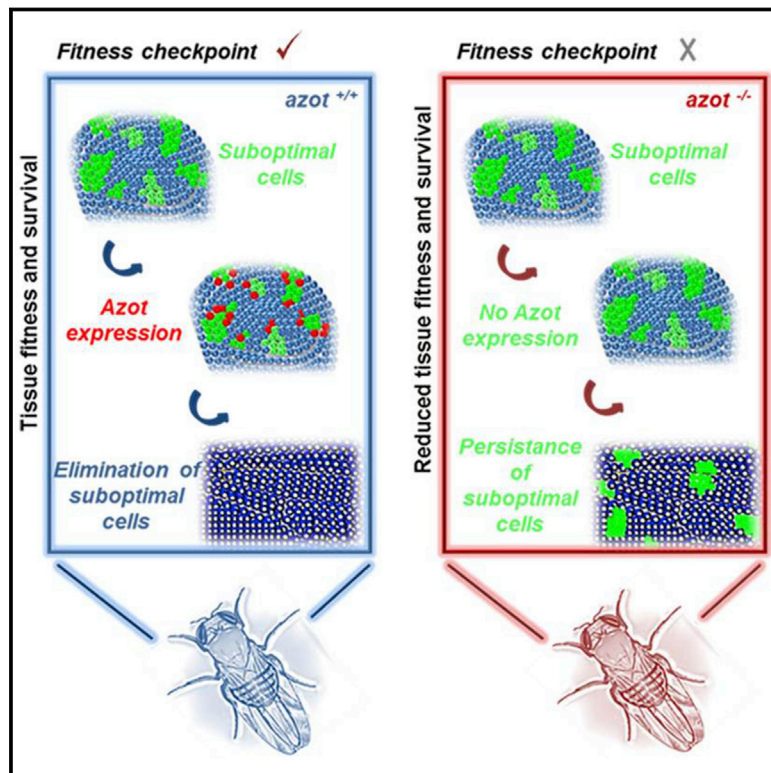
- Angelberger, S., Reinisch, W., Makristathis, A., Lichtenberger, C., Dejaco, C., Papay, P., Novacek, G., Trauner, M., Loy, A., and Berry, D. (2013). Temporal bacterial community dynamics vary among ulcerative colitis patients after fecal microbiota transplantation. *Am. J. Gastroenterol.* 108, 1620–1630.
- Barr, J.J., Auro, R., Furlan, M., Whiteson, K.L., Erb, M.L., Pogliano, J., Stotland, A., Wolkowicz, R., Cutting, A.S., Doran, K.S., et al. (2013). Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proc. Natl. Acad. Sci. USA* 110, 10771–10776.
- Basic, M., Keubler, L.M., Buettner, M., Achard, M., Breves, G., Schröder, B., Smoczek, A., Jörns, A., Wedekind, D., Zschemisch, N.H., et al. (2014).

- Norovirus triggered microbiota-driven mucosal inflammation in interleukin 10-deficient mice. *Inflamm. Bowel Dis.* 20, 431–443.
- Belkaid, Y., and Hand, T.W. (2014). Role of the microbiota in immunity and inflammation. *Cell* 157, 121–141.
- Breitbart, M., Hewson, I., Felts, B., Mahaffy, J.M., Nulton, J., Salamon, P., and Rohwer, F. (2003). Metagenomic analyses of an uncultured viral community from human feces. *J. Bacteriol.* 185, 6220–6223.
- Brüssow, H., Canchaya, C., and Hardt, W.-D. (2004). Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol. Mol. Biol. Rev.* 68, 560–602.
- Butterworth, A.D., Thomas, A.G., and Akobeng, A.K. (2008). Probiotics for induction of remission in Crohn's disease. *Cochrane Database Syst. Rev.*, CD006634.
- Cadwell, K., Patel, K.K., Maloney, N.S., Liu, T.C., Ng, A.C., Storer, C.E., Head, R.D., Xavier, R., Stappenbeck, T.S., and Virgin, H.W. (2010). Virus-plus-susceptibility gene interaction determines Crohn's disease gene Atg16L1 phenotypes in intestine. *Cell* 141, 1135–1145.
- Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336.
- Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., Fierer, N., and Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. USA* 108 (1), 4516–4522.
- Clokier, M.R., Millard, A.D., Letarov, A.V., and Heaphy, S. (2011). Phages in nature. *Bacteriophage* 1, 31–45.
- Duerkop, B.A., and Hooper, L.V. (2013). Resident viruses and their interactions with the immune system. *Nat. Immunol.* 14, 654–659.
- Duerkop, B.A., Clements, C.V., Rollins, D., Rodrigues, J.L.M., and Hooper, L.V. (2012). A composite bacteriophage alters colonization by an intestinal commensal bacterium. *Proc. Natl. Acad. Sci. USA* 109, 17621–17626.
- Dutilh, B.E., Cassman, N., McNair, K., Sanchez, S.E., Silva, G.G.Z., Boling, L., Barr, J.J., Speth, D.R., Seguritan, V., Aziz, R.K., et al. (2014). A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat. Commun.* 5, 4498.
- Eriksson, F., Tsagotis, P., Lundberg, K., Parsa, R., Mangsbo, S.M., Persson, M.A.A., Harris, R.A., and Pisa, P. (2009). Tumor-specific bacteriophages induce tumor destruction through activation of tumor-associated macrophages. *J. Immunol.* 182, 3105–3111.
- Finkbeiner, S.R., Allred, A.F., Tarr, P.I., Klein, E.J., Kirkwood, C.D., and Wang, D. (2008). Metagenomic analysis of human diarrhea: viral detection and discovery. *PLoS Pathog.* 4, e1000011.
- Fortier, L.-C., and Sekulovic, O. (2013). Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence* 4, 354–365.
- Gevers, D., Kugathasan, S., Denson, L.A., Vázquez-Baeza, Y., Van Treuren, W., Ren, B., Schwager, E., Knights, D., Song, S.J., Yassour, M., et al. (2014). The treatment-naïve microbiome in new-onset Crohn's disease. *Cell Host Microbe* 15, 382–392.
- Górski, A., Wazna, E., Dąbrowska, B.-W., Dąbrowska, K., Swiata-Jeleń, K., and Miedzybrodzki, R. (2006). Bacteriophage translocation. *FEMS Immunol. Med. Microbiol.* 46, 313–319.
- Handley, S.A., Thackray, L.B., Zhao, G., Presti, R., Miller, A.D., Droit, L., Abbink, P., Maxfield, L.F., Kambal, A., Duan, E., et al. (2012). Pathogenic simian immunodeficiency virus infection is associated with expansion of the enteric virome. *Cell* 151, 253–266.
- Hollander, D., Vadheim, C.M., Brettholz, E., Petersen, G.M., Delahunty, T., and Rotter, J.I. (1986). Increased intestinal permeability in patients with Crohn's disease and their relatives. A possible etiologic factor. *Ann. Intern. Med.* 105, 883–885.
- Huson, D.H., Mitra, S., Ruscheweyh, H.-J., Weber, N., and Schuster, S.C. (2011). Integrative analysis of environmental sequences using MEGAN4. *Genome Res.* 21, 1552–1560.
- Irving, P.M., and Gibson, P.R. (2008). Infections and IBD. *Nat. Clin. Pract. Gastroenterol. Hepatol.* 5, 18–27.
- Joossens, M., Huys, G., Cnockaert, M., De Preter, V., Verbeke, K., Rutgeerts, P., Vandamme, P., and Vermeire, S. (2011). Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut* 60, 631–637.
- Kostic, A.D., Xavier, R.J., and Gevers, D. (2014). The microbiome in inflammatory bowel disease: current status and the future ahead. *Gastroenterology* 146, 1489–1499.
- Kump, P.K., Gröchenig, H.-P., Lackner, S., Trajanoski, S., Reicht, G., Hoffmann, K.M., Deutschmann, A., Wenzl, H.H., Petritsch, W., Krejs, G.J., et al. (2013). Alteration of intestinal dysbiosis by fecal microbiota transplantation does not induce remission in patients with chronic active ulcerative colitis. *Inflamm. Bowel Dis.* 19, 2155–2165.
- Larsen, N., Vogensen, F.K., van den Berg, F.W., Nielsen, D.S., Andreasen, A.S., Pedersen, B.K., Al-Soud, W.A., Sørensen, S.J., Hansen, L.H., and Jakobsen, M. (2010). Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PLoS ONE* 5, e9085.
- Lax, S., Smith, D.P., Hampton-Marcell, J., Owens, S.M., Handley, K.M., Scott, N.M., Gibbons, S.M., Larsen, P., Shogan, B.D., Weiss, S., et al. (2014). Longitudinal analysis of microbial interaction between humans and the indoor environment. *Science* 345, 1048–1052.
- Lepage, P., Colombet, J., Marteau, P., Sime-Ngando, T., Doré, J., and Leclerc, M. (2008). Dysbiosis in inflammatory bowel disease: a role for bacteriophages? *Gut* 57, 424–425.
- Ley, R.E., Bäckhed, F., Turnbaugh, P., Lozupone, C.A., Knight, R.D., and Gordon, J.I. (2005). Obesity alters gut microbial ecology. *Proc. Natl. Acad. Sci. USA* 102, 11070–11075.
- Lindsay, J.A., Ruzin, A., Ross, H.F., Kurepina, N., and Novick, R.P. (1998). The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus*. *Mol. Microbiol.* 29, 527–543.
- Maiques, E., Ubeda, C., Campoy, S., Salvador, N., Lasa, I., Novick, R.P., Barbé, J., and Penadés, J.R. (2006). beta-lactam antibiotics induce the SOS response and horizontal transfer of virulence factors in *Staphylococcus aureus*. *J. Bacteriol.* 188, 2726–2729.
- McArdle, B.H., and Anderson, M.J. (2001). Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* 82, 290–297.
- Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S.A., Wu, G.D., Lewis, J.D., and Bushman, F.D. (2011). The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res.* 21, 1616–1625.
- Minot, S., Grunberg, S., Wu, G.D., Lewis, J.D., and Bushman, F.D. (2012). Hypervariable loci in the human gut virome. *Proc. Natl. Acad. Sci. USA* 109, 3962–3966.
- Minot, S., Bryson, A., Chehoud, C., Wu, G.D., Lewis, J.D., and Bushman, F.D. (2013). Rapid evolution of the human gut virome. *Proc. Natl. Acad. Sci. USA* 110, 12450–12455.
- Morgan, X.C., Tickle, T.L., Sokol, H., Gevers, D., Devaney, K.L., Ward, D.V., Reyes, J.A., Shah, S.A., LeLeiko, N., Snapper, S.B., et al. (2012). Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol.* 13, R79.
- Naidoo, K., Gordon, M., Fagbemi, A.O., Thomas, A.G., and Akobeng, A.K. (2011). Probiotics for maintenance of remission in ulcerative colitis. *Cochrane Database Syst. Rev.*, CD007443.
- Nicholson, J.K., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., Jia, W., and Pettersson, S. (2012). Host-gut microbiota metabolic interactions. *Science* 336, 1262–1267.
- Norman, J.M., Handley, S.A., and Virgin, H.W. (2014). Kingdom-agnostic metagenomics and the importance of complete characterization of enteric microbial communities. *Gastroenterology* 146, 1459–1469.
- Parent, K., and Wilson, I.D. (1971). Mycobacteriophage in Crohn's disease. *Gut* 12, 1019–1020.

- Peng, Y., Leung, H.C.M., Yiu, S.M., and Chin, F.Y.L. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28, 1420–1428.
- Pérez-Brocal, V., García-López, R., Vázquez-Castellanos, J.F., Nos, P., Beltrán, B., Latorre, A., and Moya, A. (2013). Study of the viral and microbial communities associated with Crohn's disease: a metagenomic approach. *Clin. Transl. Gastroenterol.* 4, e36.
- Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., Liang, S., Zhang, W., Guan, Y., Shen, D., et al. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 490, 55–60.
- Reyes, A., Haynes, M., Hanson, N., Angly, F.E., Heath, A.C., Rohwer, F., and Gordon, J.I. (2010). Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* 466, 334–338.
- Reyes, A., Semenkovich, N.P., Whiteson, K., Rohwer, F., and Gordon, J.I. (2012). Going viral: next-generation sequencing applied to phage populations in the human gut. *Nat. Rev. Microbiol.* 10, 607–617.
- Reyes, A., Wu, M., McNulty, N.P., Rohwer, F.L., and Gordon, J.I. (2013). Gnotobiotic mouse model of phage-bacterial host dynamics in the human gut. *Proc. Natl. Acad. Sci. USA* 110, 20236–20241.
- Shanahan, F., and Quigley, E.M.M. (2014). Manipulation of the microbiota for treatment of IBS and IBD-challenges and controversies. *Gastroenterology* 146, 1554–1563.
- Shen, X.J., Rawls, J.F., Randall, T., Burcal, L., Mpande, C.N., Jenkins, N., Jovov, B., Abdo, Z., Sandler, R.S., and Keku, T.O. (2010). Molecular characterization of mucosal adherent bacteria and associations with colorectal adenomas. *Gut Microbes* 1, 138–147.
- Stappenbeck, T.S., Rioux, J.D., Mizoguchi, A., Saitoh, T., Huett, A., Darfeuille-Michaud, A., Wileman, T., Mizushima, N., Carding, S., Akira, S., et al. (2011). Crohn disease: a current perspective on genetics, autophagy and immunity. *Autophagy* 7, 355–374.
- Sun, L., Nava, G.M., and Stappenbeck, T.S. (2011). Host genetic susceptibility, dysbiosis, and viral triggers in inflammatory bowel disease. *Curr. Opin. Gastroenterol.* 27, 321–327.
- Thurber, R.V., Haynes, M., Breitbart, M., Wegley, L., and Rohwer, F. (2009). Laboratory procedures to generate viral metagenomes. *Nat. Protoc.* 4, 470–483.
- Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R., and Gordon, J.I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 444, 1027–1031.
- Turnbaugh, P.J., Hamady, M., Yatsunenkov, T., Cantarel, B.L., Duncan, A., Ley, R.E., Sogin, M.L., Jones, W.J., Roe, B.A., Affourtit, J.P., et al. (2009). A core gut microbiome in obese and lean twins. *Nature* 457, 480–484.
- Uhr, J.W., Dancis, J., Franklin, E.C., Finkelstein, M.S., and Lewis, E.W. (1962). The antibody response to bacteriophage phi-X 174 in newborn premature infants. *J. Clin. Invest.* 41, 1509–1513.
- van Nood, E., Vrieze, A., Nieuwdorp, M., Fuentes, S., Zoetendal, E.G., de Vos, W.M., Visser, C.E., Kuijper, E.J., Bartelsman, J.F.W.M., Tijssen, J.G.P., et al. (2013). Duodenal infusion of donor feces for recurrent *Clostridium difficile*. *N. Engl. J. Med.* 368, 407–415.
- Virgin, H.W. (2014). The virome in mammalian physiology and disease. *Cell* 157, 142–150.
- Wagner, J., Maksimovic, J., Farries, G., Sim, W.H., Bishop, R.F., Cameron, D.J., Catto-Smith, A.G., and Kirkwood, C.D. (2013). Bacteriophages in gut samples from pediatric Crohn's disease patients: metagenomic analysis using 454 pyrosequencing. *Inflamm. Bowel Dis.* 19, 1598–1608.
- Waller, A.S., Yamada, T., Kristensen, D.M., Kultima, J.R., Sunagawa, S., Koonin, E.V., and Bork, P. (2014). Classification and quantification of bacteriophage taxa in human gut metagenomes. *ISME J.* 8, 1391–1402.
- Weitz, J.S., and Wilhelm, S.W. (2012). Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 Biol. Rep.* 4, 17.
- Willner, D., Furlan, M., Haynes, M., Schmieder, R., Angly, F.E., Silva, J., Tammadoni, S., Nosrat, B., Conrad, D., and Rohwer, F. (2009). Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. *PLoS ONE* 4, e7370.
- Willner, D., Haynes, M.R., Furlan, M., Hanson, N., Kirby, B., Lim, Y.W., Rainey, P.B., Schmieder, R., Youle, M., Conrad, D., and Rohwer, F. (2012). Case studies of the spatial heterogeneity of DNA viruses in the cystic fibrosis lung. *Am. J. Respir. Cell Mol. Biol.* 46, 127–131.
- Zackular, J.P., Baxter, N.T., Iverson, K.D., Sadler, W.D., Petrosino, J.F., Chen, G.Y., and Schloss, P.D. (2013). The gut microbiome modulates colon tumorigenesis. *mBiol.* 4.
- Zhang, Y., and LeJeune, J.T. (2008). Transduction of bla(CMY-2), tet(A), and tet(B) from *Salmonella enterica* subspecies *enterica* serovar Heidelberg to *S. Typhimurium*. *Vet. Microbiol.* 129, 418–425.
- Zhang, X., McDaniel, A.D., Wolf, L.E., Keusch, G.T., Waldor, M.K., and Acheson, D.W. (2000). Quinolone antibiotics induce Shiga toxin-encoding bacteriophages, toxin production, and death in mice. *J. Infect. Dis.* 181, 664–670.

Elimination of Unfit Cells Maintains Tissue Health and Prolongs Lifespan

Graphical Abstract



Authors

Marisa M. Merino, Christa Rhiner, ..., Barbara Hauert, Eduardo Moreno

Correspondence

emoreno@izb.unibe.ch

In Brief

Elimination of less fit cells ensured by the expression of a cell-fitness checkpoint, Azot, is important for maintaining tissue health and prolonging lifespan in *Drosophila*.

Highlights

- Fitness-based cell culling maintains tissue health
- Azot ensures the elimination of less fit cells
- Lack of azot accelerates tissue degeneration
- Improving the efficiency of cell selection extends lifespan



Elimination of Unfit Cells Maintains Tissue Health and Prolongs Lifespan

Marisa M. Merino,¹ Christa Rhiner,¹ Jesus M. Lopez-Gay,^{1,2} David Buechel,¹ Barbara Hauert,¹ and Eduardo Moreno^{1,*}

¹Institute of Cell Biology, IZB, University of Bern, Bern 3012, Switzerland

²Polarity Division and Morphogenesis, Institut Curie, CNRS UMR 3215, INSERM U934 Paris, France

*Correspondence: emoreno@izb.unibe.ch

<http://dx.doi.org/10.1016/j.cell.2014.12.017>

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/3.0/>).

SUMMARY

Viable yet damaged cells can accumulate during development and aging. Although eliminating those cells may benefit organ function, identification of this less fit cell population remains challenging. Previously, we identified a molecular mechanism, based on “fitness fingerprints” displayed on cell membranes, which allows direct fitness comparison among cells in *Drosophila*. Here, we study the physiological consequences of efficient cell selection for the whole organism. We find that fitness-based cell culling is naturally used to maintain tissue health, delay aging, and extend lifespan in *Drosophila*. We identify a gene, *azot*, which ensures the elimination of less fit cells. Lack of *azot* increases morphological malformations and susceptibility to random mutations and accelerates tissue degeneration. On the contrary, improving the efficiency of cell selection is beneficial for tissue health and extends lifespan.

INTRODUCTION

Individual cells can suffer insults that affect their normal functioning, a situation often aggravated by exposure to external damaging agents. A fraction of damaged cells will critically lose their ability to live, but a different subset of cells may be more difficult to identify and eliminate: viable but suboptimal cells that, if unnoticed, may adversely affect the whole organism (Moskalev et al., 2013).

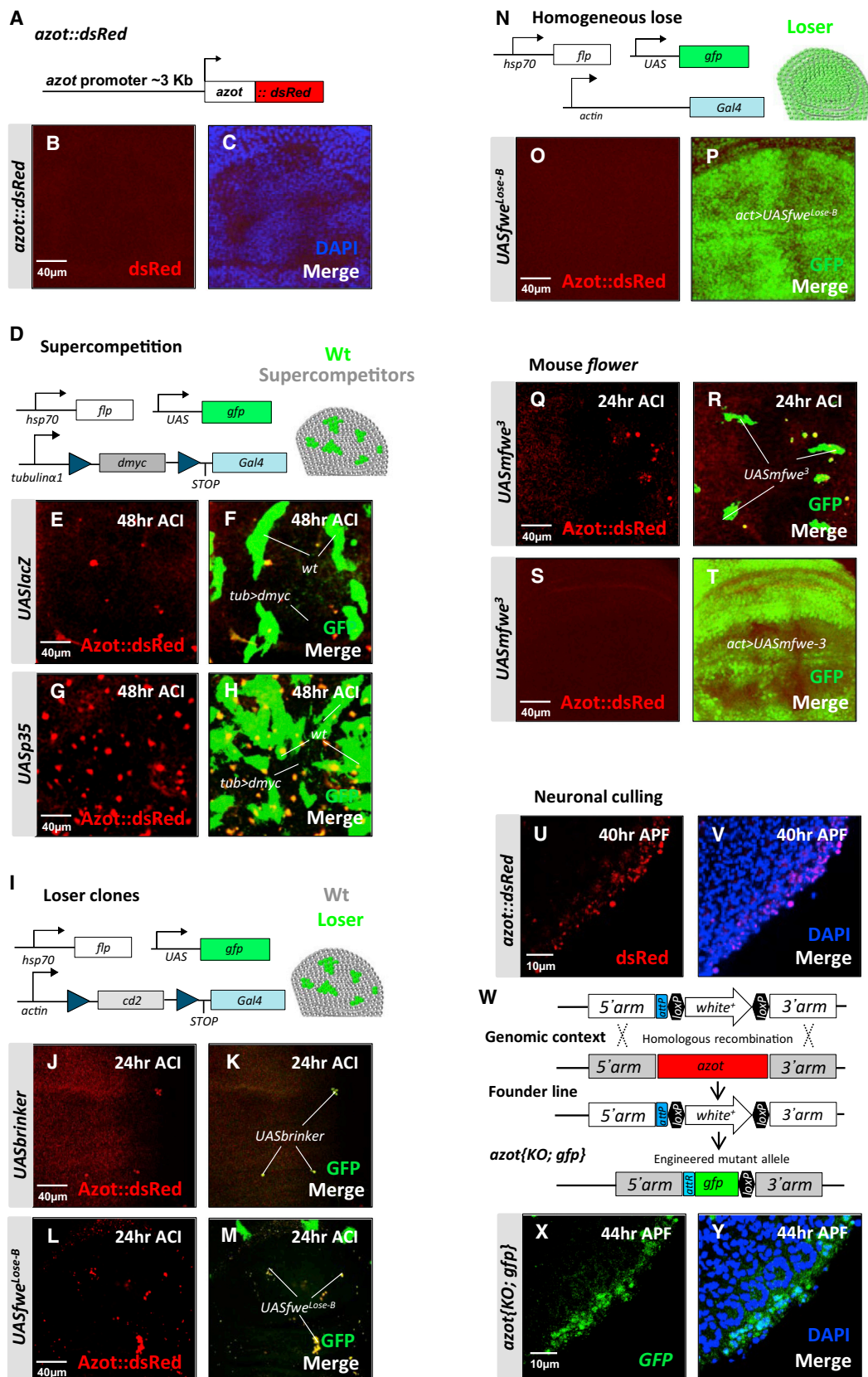
What is the evidence that viable but damaged cells accumulate within tissues? The somatic mutation theory of aging (Kennedy et al., 2012; Szilard, 1959) proposes that over time cells suffer insults that affect their fitness, for example, diminishing their proliferation and growth rates, or forming deficient structures and connections. This creates increasingly heterogeneous and dysfunctional cell populations disturbing tissue and organ function (Moskalev et al., 2013). Once organ function falls below a critical threshold, the individual dies. The theory is supported by the experimental finding that clonal mosaicism occurs at unexpectedly high frequency in human tissues as a function of time, not only in adults due to aging (Jacobs et al., 2012; Laurie et al., 2012), but also in human embryos (Vanneste et al., 2009).

Does the high prevalence of mosaicism in our tissues mean that it is impossible to recognize and eliminate cells with subtle mutations and that suboptimal cells are bound to accumulate within organs? Or, on the contrary, can animal bodies identify and get rid of unfit viable cells?

One indirect mode through which suboptimal cells could be eliminated is proposed by the “trophic theory” (Levi-Montalcini, 1987; Moreno, 2014; Raff, 1992; Simi and Ibáñez, 2010), which suggested that Darwinian-like competition among cells for limiting amounts of survival-promoting factors will lead to removal of less fit cells. However, it is apparent from recent work that trophic theories are not sufficient to explain fitness-based cell selection, because there are direct mechanisms that allow cells to exchange “cell-fitness” information at the local multicellular level (Moreno and Rhiner, 2014).

In *Drosophila*, cells can compare their fitness using different isoforms of the transmembrane protein Flower. The “fitness fingerprints” are therefore defined as combinations of Flower isoforms present at the cell membrane that reveal optimal or reduced fitness (Merino et al., 2013; Rhiner et al., 2010). The isoforms that indicate reduced fitness have been called Flower^{Lose} isoforms, because they are expressed in cells marked to be eliminated by apoptosis called “Loser cells” (Rhiner et al., 2010). However, the presence of Flower^{Lose} isoforms at the cell membrane of a particular cell does not imply that the cell will be culled, because at least two other parameters are taken into account: (1) the levels of Flower^{Lose} isoforms in neighboring cells: if neighboring cells have similar levels of Lose isoforms, no cell will be killed (Merino et al., 2013; Rhiner et al., 2010); (2) the levels of a secreted protein called Sparc, the homolog of the Sparc/Osteonectin protein family, which counteracts the action of the Lose isoforms (Portela et al., 2010).

Remarkably, the levels of Flower isoforms and Sparc can be altered by various insults in several cell types, including: (1) the appearance of slowly proliferating cells due to partial loss of ribosomal proteins, a phenomenon known as cell competition (Bailon and Basler, 2014; de Beco et al., 2012; Hogan et al., 2011; Morata and Ripoll, 1975; Moreno et al., 2002; Tamori and Deng, 2011); (2) the interaction between cells with slightly higher levels of d-Myc and normal cells, a process termed supercompetition (de la Cova et al., 2004; Moreno and Basler, 2004); (3) mutations in signal transduction pathways like Dpp signaling (Portela et al., 2010; Rhiner et al., 2010); or (4) viable neurons forming part of incomplete ommatidia (Merino et al., 2013). Intriguingly, the role of Flower isoforms is cell type specific, because certain isoforms acting as Lose marks in epithelial cells



(legend on next page)

(Rhiner et al., 2010) are part of the fitness fingerprint of healthy neurons (Merino et al., 2013). Therefore, an exciting picture starts to appear, in which varying levels of Sparc and different isoforms of Flower are produced by many cell types, acting as direct molecular determinants of cell fitness.

Here, we aimed to clarify how cells integrate fitness information in order to identify and eliminate suboptimal cells. Subsequently, we analyzed what are the physiological consequences of efficient cell selection for the whole organism.

RESULTS

Azot Is Expressed in Cells Undergoing Negative Selection

In order to discover the molecular mechanisms underlying cell selection in *Drosophila*, we analyzed genes transcriptionally induced using an assay where WT cells (*tub>Gal4*) are outcompeted by dMyc-overexpressing supercompetitors (*tub>dmyc*) (Figure 1D) due to the increased fitness of these dMyc-overexpressing cells (Rhiner et al., 2010). The expression of CG11165 (Figure S1A available online) was strongly induced 24 hr (hr) after the peak of *flower* and *sparc* expression (Figure S1B). In situ hybridization revealed that CG11165 mRNA was specifically detected in Loser cells that were going to be eliminated from wing imaginal discs due to cell competition (Figure S1C). The gene, which we named *ahuizotl* (*azot*) after a multihanded Aztec creature selectively targeting fishing boats to protect lakes (Reeves, 2006), consists of one exon. *azot*'s single exon encodes for a four EF-hand-containing cytoplasmic protein of the canonical family (Figures S1D and S1E) that is conserved, but uncharacterized, in multicellular animals (Figure S1A).

To monitor Azot expression, we designed a translational reporter resulting in the expression of Azot::dsRed under the control of the endogenous *azot* promoter in transgenic flies (Figure 1A). Azot expression was not detectable in most wing imaginal discs under physiological conditions in the absence of competition (Figures 1B and 1C). We next generated mosaic tissue of two clonal populations, which are known to trigger competitive interactions resulting in elimination of otherwise viable cells. Cells with lower fitness were created by confronting WT cells with dMyc-overexpressing cells (Figures 1E–1H) (Moreno and Basler, 2004), by downregulating Dpp signaling (Moreno et al., 2002) (Figures 1I–1K), by overexpressing Flower^{Lose} isoforms (Rhiner et al., 2010) (Figures 1L and 1M), in cells with reduced Wg signaling (Figure S1F) (Vincent et al., 2011), by suppressing Jak-Stat signaling (Rodrigues et al., 2012) in subgroups of cells (Figure S1G) or by generating *Minute* clones (Lolo et al., 2012; Morata and Ripoll,

1975; Simpson, 1979) (Figure S1H). Azot expression was not detectable in nonmosaic tissue of identical genotype (Figures 1N–1P; Figures S1I and S1J), nor in control clones overexpressing *UASlacZ* (Figure S1K). On the contrary, Azot was specifically activated in all tested scenarios of cell competition, specifically in the cells undergoing negative selection (“Loser cells”) (Figures 1D–1M). Azot expression was not repressed by the caspase inhibitor protein P35 (Figures 1G and 1H).

Because Flower proteins are conserved in mammals (Petrova et al., 2012), we decided to test if they are also able to regulate *azot*. Mouse Flower isoform 3 (mFlower³) has been shown to act as a “classical” Lose isoform, driving cell elimination when expressed in scattered groups of cells (Petrova et al., 2012), a situation where *azot* was induced in Loser cells (Figures 1Q and 1R) but is not inducing cell selection when expressed ubiquitously a scenario where *azot* was not expressed (Figures 1S and 1T). This shows that the mouse Flower^{Lose} isoforms function in *Drosophila* similarly to their fly homologs.

Interestingly, *azot* is not a general apoptosis-activated gene because its expression is not induced upon *eiger*, *hid*, or *bax* activation, which trigger cell death (Fuchs and Steller, 2011; Gaumer et al., 2000) (Figures S1L–S1N). Azot was also not expressed during elimination of cells with defects in apicobasal polarity (Figure S1O) or undergoing epithelial exclusion-mediated apoptosis (*dCsk*) (Figures S1P and S1Q) (Vidal et al., 2006).

Next, we analyzed if *azot* is expressed during the elimination of peripheral photoreceptors in the pupal retina, a process mediated by Flower-encoded fitness fingerprints (Merino et al., 2013). Thirty-six to 38 hr after pupal formation (APF), when Flower^{Lose-B} expression begins in peripheral neurons (Merino et al., 2013), we could not detect Azot expression in the peripheral edge (Figures S1R–S1U). At later time points (40 and 44 hr APF), Azot expression is visible and restricted to the peripheral edge where photoreceptor neurons are eliminated (Figures 1U and 1V). This expression was confirmed with another reporter line, *azot{KO; gfp}*, where *gfp* was directly inserted at the *azot* locus using genomic engineering techniques (Huang et al., 2009) (Figures 1W–1Y).

From these results, we conclude that Azot expression is activated in several contexts where suboptimal and viable cells are normally recognized and eliminated.

Azot Is Required to Eliminate Loser Cells and Unwanted Neurons

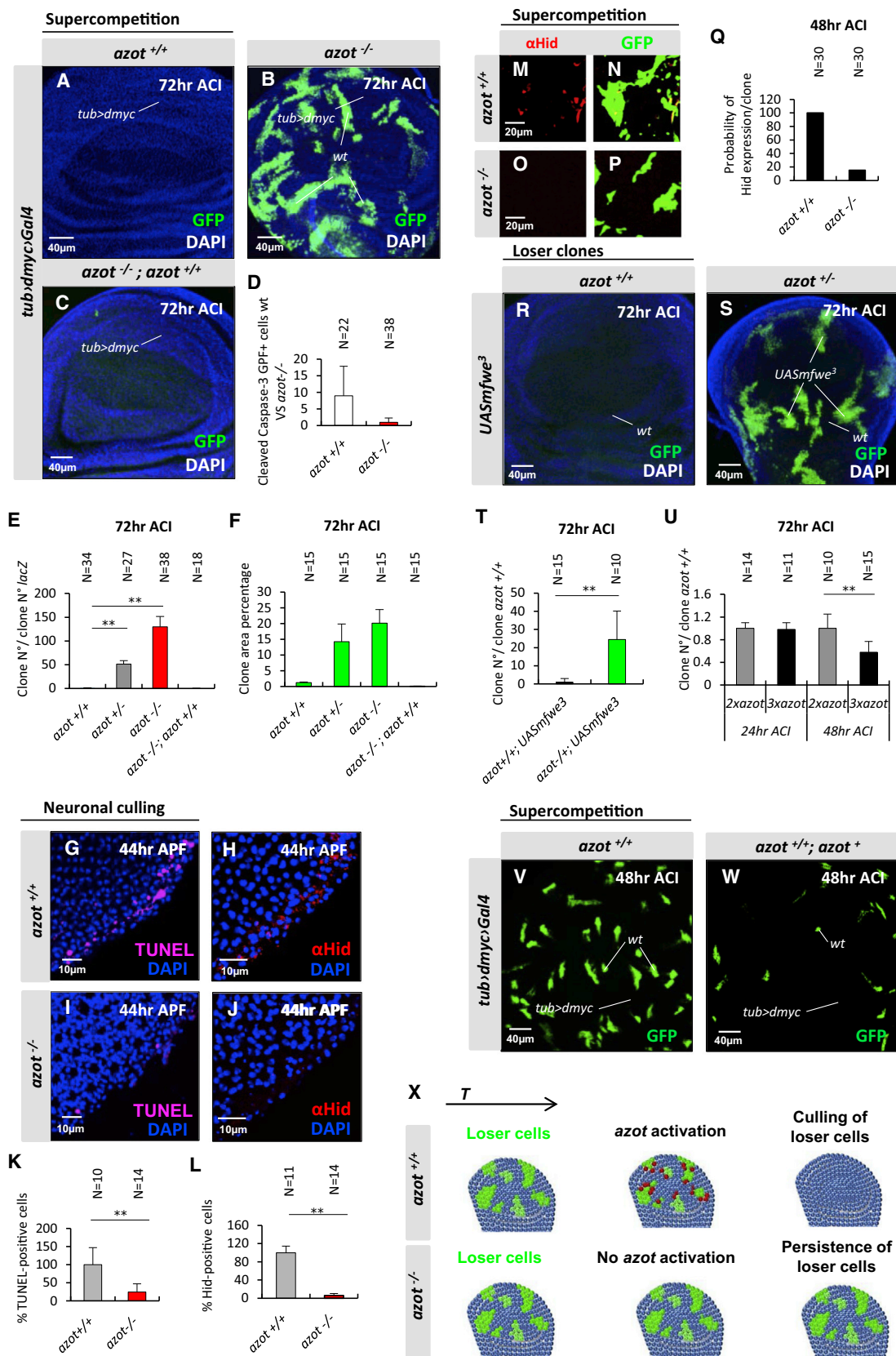
To understand Azot function in cell elimination, we generated *azot* knockout (KO) flies, whereby the entire *azot* gene was deleted (Figure 1W). Next, we analyzed Azot function using

Figure 1. Azot Is Expressed during Cell Selection of Viable Unfit Cells

(A–M) Expression analysis of Azot during different types of cell competition. For all pictures, Azot::dsRed reporter (A) is in red, and merges show outcompeted clones (green, marked with GFP) of several genotypes. DAPI is in blue. The following genotypes were analyzed: (B and C) *azot::dsRed* and (D–F) *tub>dmyc* background (black) and WT cells marked with GFP (green). Clones were generated as shown in (D) and analyzed 48 hr ACI. (G and H) *tub>dmyc* background (black) and WT cells marked with GFP (green) expressing in addition to the P35 caspase inhibitor (*UASp35*). Forty-eight hourr ACI. (I–M) Flip-out clones (green) generated as shown in (I) and overexpressing *brinker* (*UASbrinker*) (J and K), *fwe*^{Lose-B} (*UASfwe*^{Lose-B}) (L and M), or *mfw*³ (*UASmfw*³). (Q and R) Twenty-four hourr ACI.

(N–P, S, and T) General overexpression of *UASfwe*^{Lose-B} and *UASmfw*³ using the actin promoter as shown in (N).

(U–Y) Pupal retinas at different developmental time points. (U and V) Expression analysis of Azot (red), using Azot::dsRed, in peripheral photoreceptors at 40 hr after pupa formation (APF) (U and V). (W) Genomic engineering strategy used for the generation of *azot* knockout (KO) flies. (X and Y) GFP expression (green) driven by the *azot* promoter in *azot{KO; gfp}*, 44 hr APF, DAPI (blue, Y).



(legend on next page)

dmyc-induced competition. In the absence of Azot function, loser cells were no longer eliminated (Figures 2A–2F), showing a dramatic 100-fold increase in the number of surviving clones (Figures 2B and 2E). Loser cells occupied more than 20% of the tissue 72 hr after clone induction (ACI) (Figures 2B and 2F). Moreover, using *azot*^{KO}; *gfp* homozygous flies (that express GFP under the *azot* promoter but lack Azot protein), we found that loser cells survived and showed accumulation of GFP (Figures S2A and S2B). From these results, we conclude that *azot* is expressed by loser cells and is essential for their elimination.

In addition, clone removal was delayed in an *azot* heterozygous background (50-fold increase, 15%) (Figures 2E and 2F), compared to control flies with normal levels of Azot (1-fold, 1%) (Figures 2A, 2E, and 2F). Cell elimination capacity was fully restored by crossing two copies of *Azot::dsRed* into the *azot*^{−/−} background (0.5-fold, 0.2%) demonstrating the functionality of the fusion protein (Figures 2C, 2E, and 2F). Silencing *azot* with two different RNAis was similarly able to halt selection during *dmyc*-induced competition (Figures S2C–S2P). Next, in order to determine the role of Azot's EF hands, we generated and overexpressed a mutated isoform of Azot (Pm4Q12) carrying, in each EF hand, a point mutation known to abolish Ca²⁺ binding (Maune et al., 1992). Although overexpression of wild-type *azot* in negatively selected cells did not rescue the elimination (Figures S2E, S2I, S2L, and S2P), overexpression of the mutant *Azot*Pm4Q12 reduced cell selection (Figures S2H, S2I, S2O, and S2P), functioning as a dominant-negative mutant. This shows that Ca²⁺ binding is important for Azot function. Finally, staining for apoptotic cells corroborated that the lack of Azot prevents cell elimination, because cell death was reduced 8-fold in mosaic epithelia containing loser cells (Figure 2D).

Next, we analyzed the role of *azot* in elimination of peripheral photoreceptor neurons in the pupal retina using homozygous *azot* KO flies (Figures 2G–2L). Pupal retinas undergoing photoreceptor culling (44 hr APF) of *azot*^{+/+} and *azot*^{−/−} flies were stained for the cell death marker TUNEL (Figures 2G and 2I) and the pro-apoptotic factor Hid (Figures 2H and 2J). Consistent with the expression pattern of Azot, the number of Hid and TUNEL-positive cells was dramatically decreased in *azot*^{−/−} retinas (Figures 2I–2L) compared to *azot*^{+/+} retinas (Figures 2G, 2H, 2K, and 2L).

Those results showed that Azot was required to induce cell death and Hid expression during neuronal culling. Therefore, we tested if that was also the case in the wing epithelia during *dmyc*-induced competition. We found that Hid was expressed

in loser cells and that the expression was strongly reduced in the absence of Azot function (Figures 2M–2Q).

Finally, forced overexpression of *Flower*^{Lose} isoforms from *Drosophila* (Figures S2Q, S2R, and S2T) and mice (Figures 2R–2T; Figures S2S and S2U) were unable to mediate WT cell elimination when Azot function was impaired by mutation or silenced by RNAi.

These results suggested that *azot* function was dose sensitive, because heterozygous *azot* mutant flies displayed delayed elimination of loser cells when compared with *azot* WT flies (Figure 2E). We therefore took advantage of our functional reporter *Azot::dsRed* (Figures 2C and 2E) to test whether cell elimination could be enhanced by increasing the number of genomic copies of *azot*. We found that tissues with three functional copies of *azot* were more efficient eliminating loser cells during *dmyc*-induced competition and most of the clones were culled 48 hr ACI (Figures 2U–2W).

From these results, we conclude that *azot* expression is required for the elimination of Loser cells and unwanted neurons (Figure 2X).

Azot Maintains Tissue Fitness during Development

Next, we asked what could be the consequences of decreased cell selection at the tissue and organismal level. To this end, we took advantage of the viability of homozygous *azot* KO flies. We observed an increase of several developmental aberrations. We focused on the wings, where cell competition is best studied and, because aberrations were easy to define, which comprised melanotic areas, blisters, and wing margin nicks (Figures 3A–3E). Wing defects of *azot* mutant flies could be rescued by introducing two copies of *azot::dsRed*, showing that the phenotypes are specifically caused by loss of Azot function (Figures 3A–3E).

Next, we reasoned that mild tissue stress should increase the need for fitness-based cell selection after damage. First, in order to generate multicellular tissues scattered with suboptimal cells, we exposed larvae to UV light (Figure 3F) and monitored Azot expression in wing discs of UV-irradiated WT larvae, which were stained for cleaved caspase-3, 24 hr after treatment (Figures 3G–3K). Under such conditions, Azot was found to be expressed in cleaved caspase-3-positive cells (Figures 3H–3K). All Azot-positive cells showed caspase activation and 17% of cleaved caspase-positive cells expressed Azot (Figure 3G). This suggested that Azot-expressing cells are culled from the tissue. To confirm this, we looked at later time points (3 days after

Figure 2. Azot Is Required to Eliminate Loser Cells and Unwanted Neurons

(A–F) Analysis of *azot* KO during *dmyc*-induced supercompetition 72 hr ACI. (D) Quantification of cleaved caspase-3 and GFP-positive cells during *dmyc*-induced supercompetition in *azot*^{+/+} and *azot*^{−/−} backgrounds ($p < 0.01$) 72 hr ACI. (E) Quantification of number of clones; the following backgrounds were analyzed: (A and E) *azot*^{+/+}, (E) *azot*^{+/−} ($p < 0.01$), (B and E) *azot*^{−/−} ($p < 0.01$), and (C and E) *azot*^{−/−}; *+* ($p > 0.05$). (F) Percentage of the wing pouch occupied by the wt cells in the (A and F) *azot*^{+/+}, (F) *azot*^{+/−}, (B and F) *azot*^{−/−}, (C and F) *azot*^{−/−}; *+*.

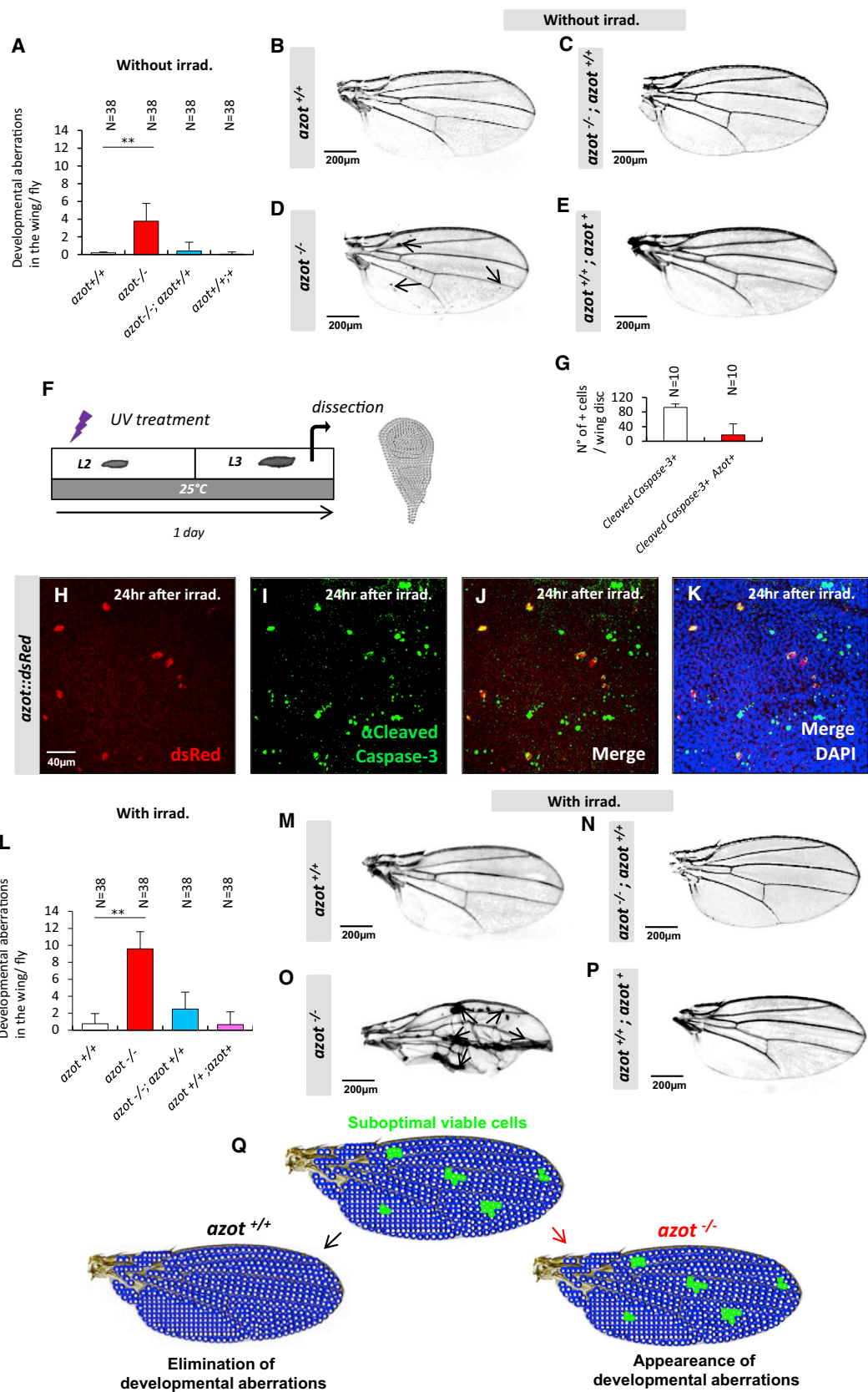
(G–L) Role of *azot* during neuronal culling in the pupal retina. (K and L) Quantification of the number of apoptotic (TUNEL-positive, magenta) or Hid-expressing (red) peripheral photoreceptors, in *azot*^{+/+} (G, H, K, and L) and *azot*^{−/−} (I, J, K, and L) flies. DAPI is in blue.

(M–Q) Hid expression (red) in loser clones (green) during supercompetition 48 hr ACI in *azot*^{+/+} (M, N, and Q) and *azot*^{−/−} (O–Q) backgrounds.

(R–T) Seventy-two hour ACI *mfwe*³-overexpressing clones (*UASmfwe*³) in *azot*^{+/+} (R and T) and *azot*^{+/−} (S and T) backgrounds ($p < 0.01$).

(U–W) Analysis of an extra genomic copy of *azot* during *dmyc*-induced supercompetition. (U) Quantification of the number of clones during *dmyc*-induced supercompetition with or without an extra genomic copy of *azot*. (V and W) Discs analyzed 48 hr ACI in *azot*^{+/+} (V) and *azot*^{+/+}; *azot*⁺ ($p < 0.01$) (W).

(X) Azot expression is required for cell-competition-mediated apoptosis of loser cells. Data are represented as mean \pm SEM.



(legend on next page)

irradiation; Figure S3A) and found that the increase in Azot-positive cells was no longer detectable (Figures S3B–S3D). The elimination of *azot*-expressing cells after UV irradiation required *azot* function, because cells revealed by reporter *azot*{*KO*; *gfp*}, that express GFP instead of Azot, persisted in wing imaginal discs from *azot*-null larvae (Figures S3E–S3G). We therefore tested if lack of *azot* leads to a faster accumulation of tissue defects during organ development upon external damage. We irradiated *azot*^{−/−} pupae 0 stage (Figures 3L–3P) and compared the number of morphological defects in adult wings to those in nonirradiated *azot* KO flies (Figures 3A–3E). We found that aberrations increased more than 2-fold when compared to nonirradiated *azot*^{−/−} flies (Figures 3L–3P).

In order to functionally discriminate whether *azot* belongs to genes regulating apoptosis in general or is dedicated to fitness-based cell selection, we examined if *azot* silencing prevented Eiger/TNF-induced cell death (*GMR-Gal4, UASeiger*) (Figures S3H–S3N). Inhibiting apoptosis (*UASp35*) or *eiger* (*UASRNAieiger*) rescued eye ablation, whereas *azot* silencing and overexpression of AzotPm4Q12 did not (Figures S3I–S3N). Furthermore, *azot* silencing did not impair apoptosis during genitalia rotation (Figures S3O–S3R) (Suzanne et al., 2010) or cell death of epithelial precursors in the retina (Figures S3S–S3V) (Wolff and Ready, 1991).

The results showed above highlight the consequences of nonfunctional cell-quality control within developing tissues (Figure 3Q).

***azot* Promoter Computes Relative Flower^{Lose} and Sparc Levels**

Next, we performed epistasis analyses to understand at which level *azot* is transcriptionally regulated. For this purpose, we again used the assay where WT cells are outcompeted by dMyc-overexpressing supercompetitors (Figure 1D). We have previously observed that *azot* induction is triggered upstream of caspase-3 activation and accumulated in outcompeted cells unable to die (Figures 1G and 1H). Then, we genetically modified upstream events of cell selection (Figures 4A–4G): silencing *fwe*^{Lose} transcripts by RNAi or overexpressing Sparc, both blocked the induction of Azot::dsRed in WT loser cells (Figures 4A–4D and 4G). In contrast, when outcompeted WT cells were additionally “weakened” by Sparc downregulation using RNAi, Azot is detected in almost all loser cells (Figures 4E–4G) compared to its more limited induction in the presence of endogenous Sparc (Figures 1E and 1F and 4G). Inhibiting JNK signaling with *UASpuc* (Martín-Blanco et al., 1998; Moreno et al., 2002) did not suppress Azot expression (Figures S4A and S4B).

Next, we analyzed the activation of Azot upon irradiation. Strikingly, we found that all Azot expression after irradiation was eliminated when Flower Lose was silenced and also when relative differences of Flower Lose were diminished by overexpressing high levels of Lose isoforms ubiquitously (Figures 4H–4K; Figure S4C). On the contrary, Azot was not suppressed after irradiation by expressing the prosurvival factor Bcl-2 or a p53 dominant negative (Brodsky et al., 2000; Gaumer et al., 2000) (Figures S4C–S4G). Those results show that Azot expression during competition and upon irradiation requires differences in Flower Lose relative levels.

Finally, we analyzed the regulation of Azot expression in neurons. Silencing *fwe* transcripts by RNAi blocked the induction of Azot::dsRed in peripheral photoreceptors (Figures 4L and 4M; Figure S4H). Because Wingless signaling induces Flower^{Lose-B} expression in peripheral photoreceptors (Merino et al., 2013), we tested if overexpression of Daxin, a negative regulator of the pathway (Willert et al., 1999), affected Azot levels and found that it completely inhibited Azot expression (Figures S4H–S4J). Similarly, overexpression of the cell competition inhibitor Sparc also fully blocked Azot endogenous expression in the retina (Figures S4H, S4K, and S4L). Finally, ectopic overexpression of Flower^{Lose-B} in scattered cells of the retina was sufficient to trigger ectopic Azot activation (Figures S4M–S4O). Those results show that photoreceptor cells also can monitor the levels of Sparc and the relative levels of Flower^{Lose-B} before triggering Azot expression (Figure S4P).

The results described above suggest that the *azot* promoter integrates fitness information from neighboring cells, acting as a relative “cell-fitness checkpoint” (Figures 4N–4Q).

Cell Selection Is Active during Adulthood

To test if fitness-based cell selection is a mechanism active not only during development, but also during adult stages, we exposed WT adult flies to UV light and monitored Azot and Flower expression in adult tissues (Figures 5A–5T). UV irradiation of adult flies triggered cytoplasmic Azot expression in several adult tissues including the gut (Figures 5B–5E; Figures S5A and S5B) (Lemaître and Miguel-Aliaga, 2013) and the adult brain (Figures 5F–5J) (Fernández-Hernández et al., 2013). Likewise, UV irradiation of adult flies triggered Flower Lose expression in the gut (Figures 5K–5N) and in the brain (Figures 5O–5T). Irradiation-induced Azot expression was unaffected by Bcl-2 but was eliminated when Flower Lose was silenced or when relative differences of Flower Lose were diminished in the gut (Figures S5C–S5E) and in the adult brain (Figures S5F–S5H). This suggests that the process of cell selection is active throughout the life history of the animal. Further confirming this conclusion,

Figure 3. *Azot* Mutants Show Developmental Aberrations

(A–E) Wings of 10- to 13-day-old flies and quantification of developmental aberrations in the wing of each genotype, ** < 0.01. (A and B) *azot*^{+/+}, (A and C) *azot*^{−/−}; *azot*^{+/+}, (A and D) *azot*^{−/−} and (A and E) *azot*^{+/+}; *azot*⁺. (F–K) Azot and cleaved caspase-3 expression upon UV irradiation (2×10^{-2} J irradiation dose during second instar larvae, treatment as shown in F). (G) Quantification of the percentage of Azot and cleaved caspase-3-expressing cells after UV irradiation. (H) Azot::dsRed expression after UV irradiation (red), (I) cleaved caspase-3 (green) after UV irradiation, (J) merge, and (K) merge with DAPI (blue). (L–P) Quantification of developmental aberrations and images of wings from 10- to 13-day-old flies after UV treatment (2×10^{-2} J, pupae stage 0) of genotypes (L and M) *azot*^{+/+}, (L and N) *azot*^{−/−}; *azot*^{+/+}, (L and O) *azot*^{−/−}, and (L and P) *azot*^{+/+}; *azot*⁺. (Q) Scheme showing the requirement of *azot* function for preventing developmental aberrations. Data are represented as mean ± SEM.

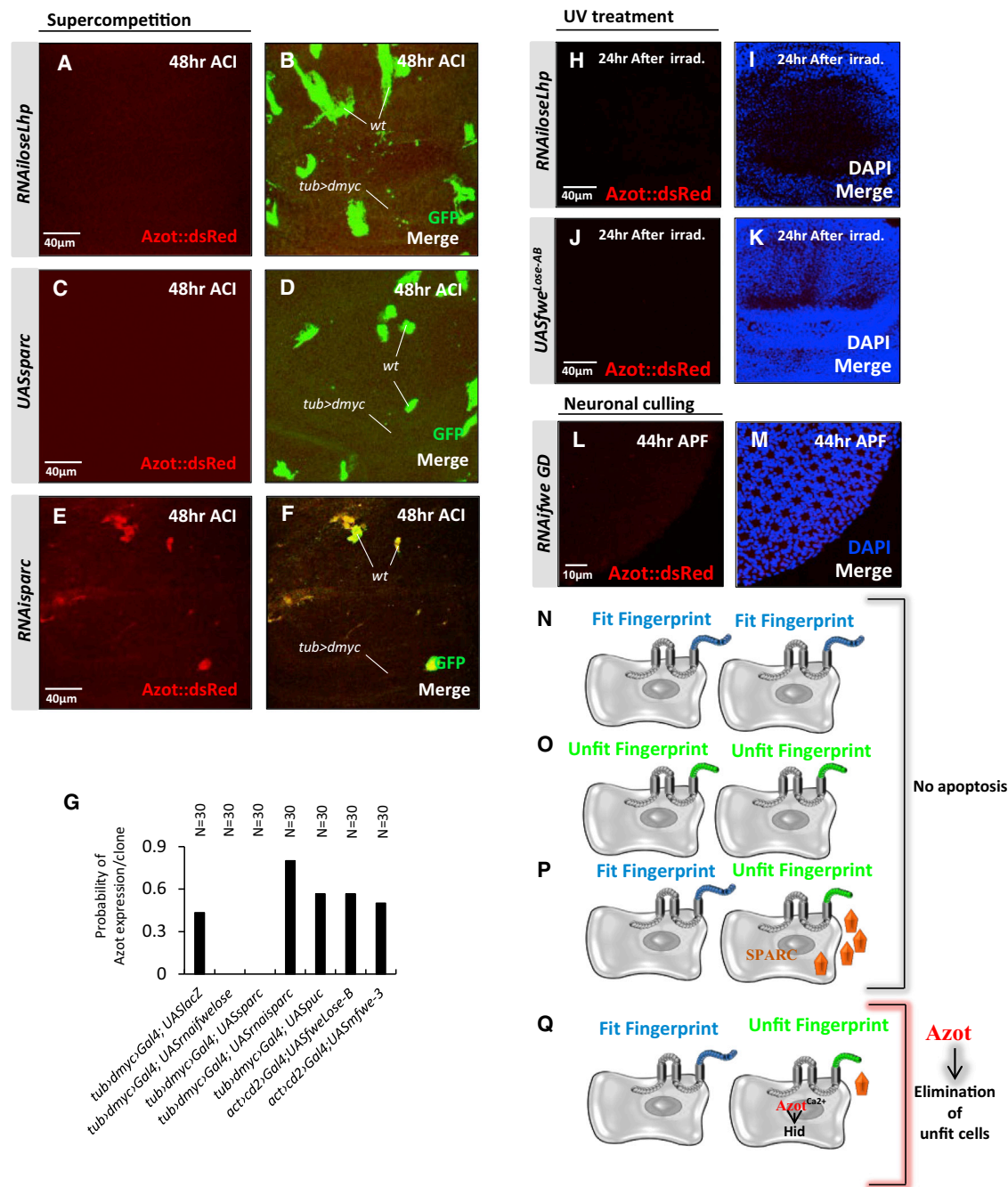


Figure 4. The *azot* Promoter Computes Relative *Flower^{Lose}* and *Sparc* Levels

(A–F) Epistasis analysis of the following genotypes during *dmyc*-induced supercompetition. (A and B) *UASRNAiifwe^{loseLhp}*, (C and D) *UASsparc*, and (E and F) *UASRNAiifwe^{loseA}*. *Azot::dsRed* is shown in red (A, C, and E) and merges with GFP in (B, D, and F).

(G) Graph showing the probability of finding *Azot* expression in a GFP marked clone in several genotypes.

(H–J) *Azot::dsRed* expression after UV irradiation (red) is suppressed when *UASRNAiifwe^{loseLhp}* (H and I) or *UASfwe^{lose-B}* and *UASfwe^{lose-A}* (J and K) are expressed ubiquitously. Quantified in Figure S4C.

(L and M) Epistasis analysis of *Azot* expression in the *Drosophila* retina. Pupal retinas dissected 44 hr APF of *GMR-Gal4; RNAiifwe (GD)*. *Azot* expression shown in red (L) and merge with nuclear marker DAPI in blue (M). Quantified in Figure S4H.

(N) *Azot* is not expressed in cells without *Flower^{Lose}* isoforms.

(O–Q) Cells expressing *Flower^{Lose}* but that are either surrounded by cells with equal or higher levels of *Flower^{Lose}* (O) or express high levels of *Sparc* (P) also do not activate *azot* expression. Cells with higher relative levels of *Lose* and not enough *Sparc* induce the expression of *azot* and are eliminated (Q).

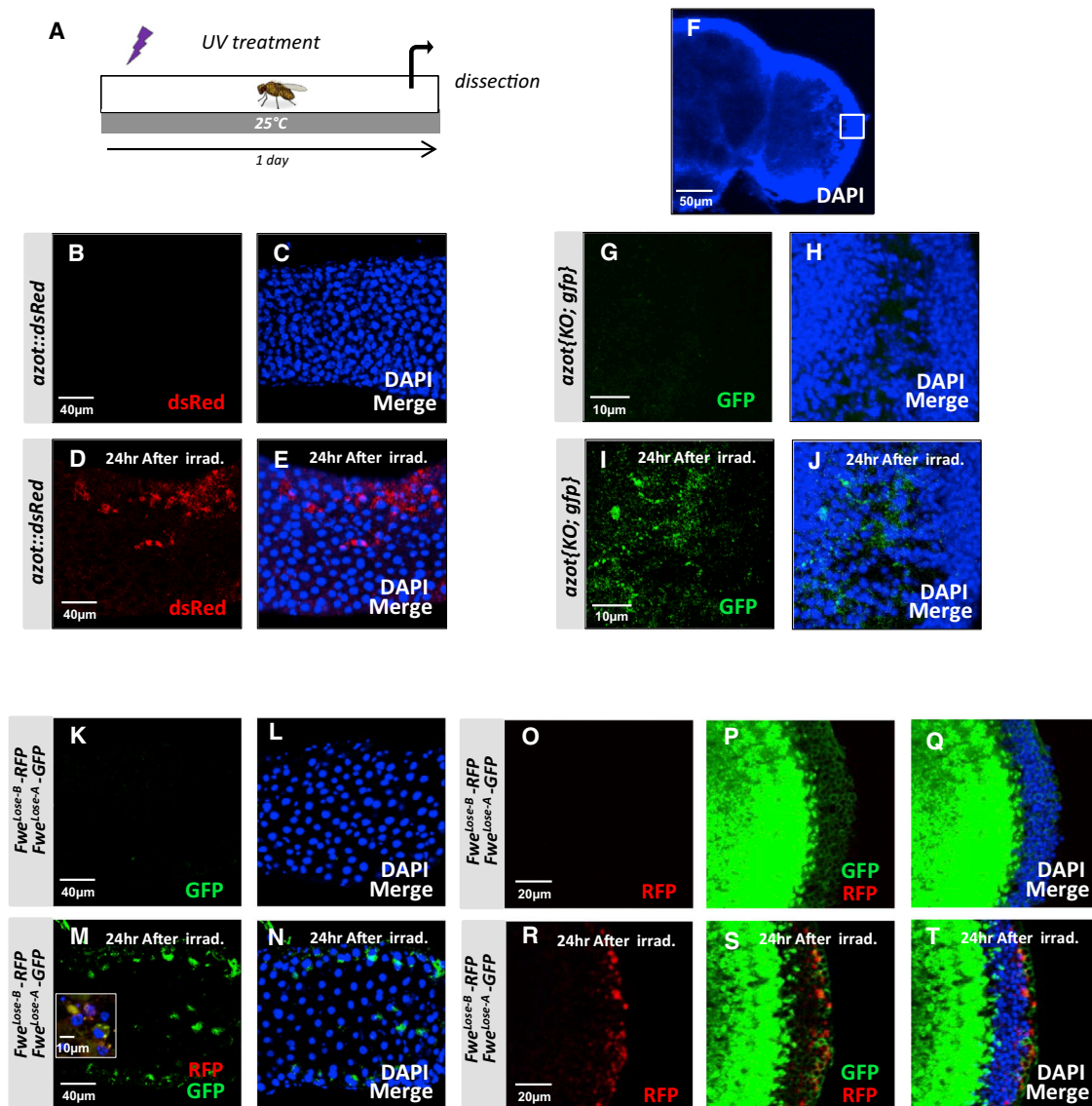


Figure 5. Expression of Flower Isoforms and Azot in Adult Flies with and without UV Irradiation

(A–E) Expression analysis of Azot (red, B and D) in the midgut without (B and C) and with (D and E) UV-irradiation treatment (as shown in A); (C) and (E) show merges with DAPI.

(F–J) Expression analysis of Azot using reporter line *azot{KO; gfp}* in the adult brain without (G and H) and after (I and J) UV-irradiation treatment merges with DAPI in (H and J).

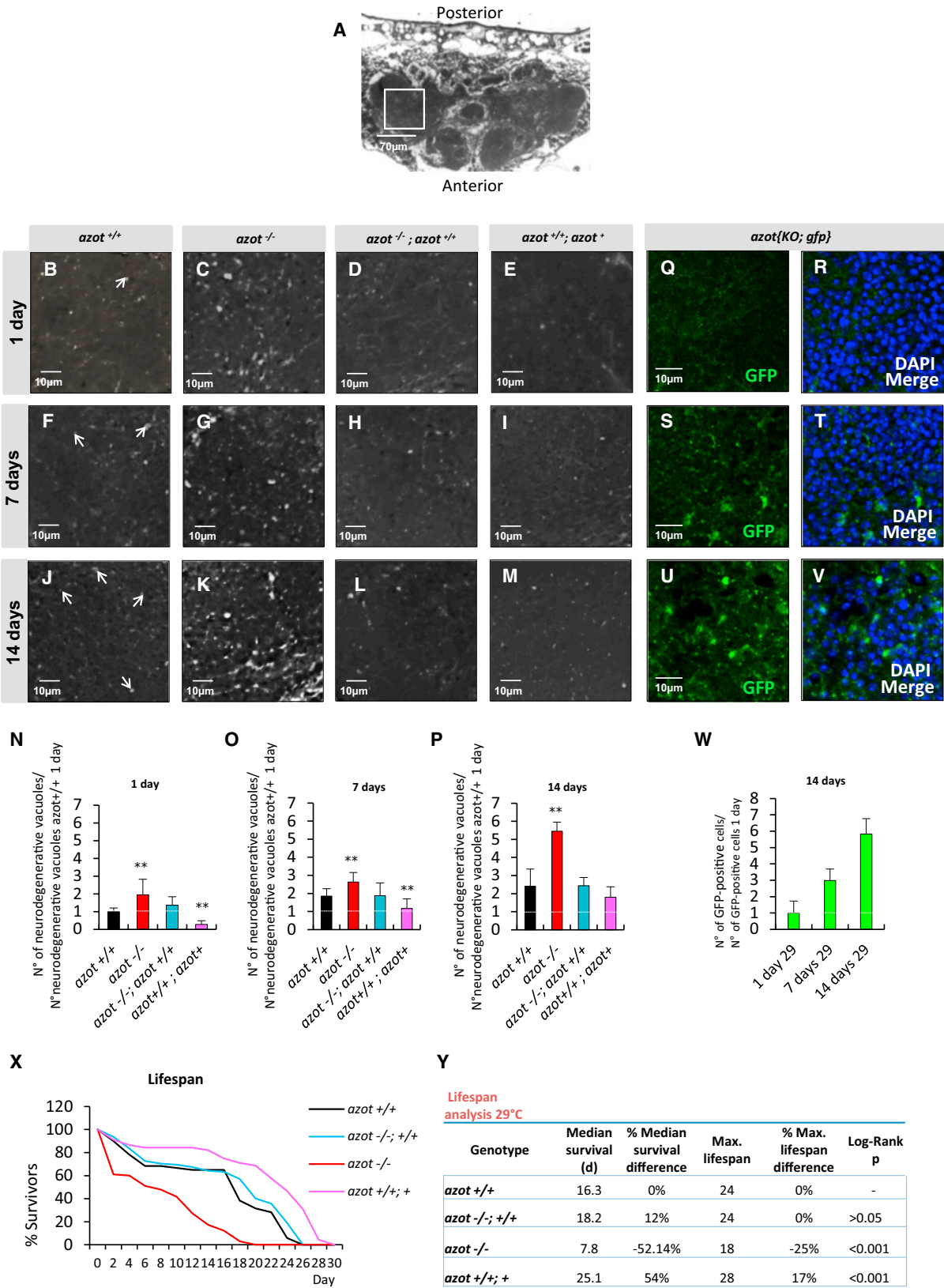
(K–T) Expression analysis of Flower Lose isoforms Lose A (green) and Lose B (red) (*flower Lose-A-GFP*, *flower Lose-B-RFP*). (K and M) In the midgut without (K and L) and with (M and N) UV-irradiation treatment. (L and N) merges with DAPI. Inset in (M) shows *Fwe^{Lose-A}* and *Fwe^{Lose-B}* expression at higher magnification. (O–T) Expression of Flower Lose isoforms in the adult brain without (O–Q) and after (R–T) UV irradiation, merges with DAPI in (Q and T).

Azot function was essential for survival after irradiation, because more than 99% of *azot* mutant adults died 6 days after irradiation, whereas only 62.4% of WT flies died after the same treatment (Figure S5I). The percentage of survival correlated with the dose of *azot* because adults with three functional copies of *azot* had higher median survival and maximum lifespan than WT flies, or null mutant flies rescued with two functional *azot* transgenes (Figure S5J).

Those results show that in adult tissues external damage can induce cell-fitness deficits.

Role of Cell Selection during Aging

Lack of cell selection could affect the whole organism by two nonexclusive mechanisms. First, the failure to detect precancerous cells, which could lead to cancer formation and death of the individual. Second, the time-dependent accumulation of unfit



(legend on next page)

but viable cells could lead to accelerated tissue and organ decay. We therefore tested both hypotheses.

It has been previously shown that cells with reduced levels for cell polarity genes like *scrib* or *dlg* are eliminated but can give rise to tumors when surviving (Igaki et al., 2009; Parisi et al., 2014; Tamori et al., 2010). We therefore checked if *azot* functions as a tumor suppressing mechanism in those cells (Figures S6A–S6M). Elimination of *dlg* and *scrib* mutant cells was not affected by RNAi against *azot* (Figures S6D–S6M) or when Azot function was impaired by mutation (Figures S6N–S6R), in agreement with the absence of *azot* induction in these mutant cells (Figures S1O and S6A–S6C). However, *azot* RNAi or the same *azot* mutant background efficiently rescued the elimination of clones with reduced Wg signaling (Vincent et al., 2011) (Figures S6J–S6M, S6Q, and S6R).

Moreover, the high number of suboptimal cells produced by UV treatment did not lead to tumoral growth in *azot*-null background (Figures S3E–S3G). Thus, tumor suppression mechanisms are not impaired in *azot* mutant backgrounds, and tumors are not more likely to arise in *azot*-null mutants.

Second, we tested whether the absence of *azot* accelerates tissue fitness decay in adult tissues. We focused on the adult brain, where neurodegenerative vacuoles develop over time and can be used as a marker of aging (Liu et al., 2012). We compared the number of vacuoles appearing in the brain of flies lacking *azot* (*azot*^{−/−}), WT flies (*azot*^{+/+}), flies with one extra genomic copy of the gene (*azot*^{+/+}; *azot*⁺), and mutant flies rescued with two genomic copies of *azot* (*azot*^{−/−}; *azot*^{+/+}). For all the genotypes analyzed, we observed a progressive increase in the number and size of vacuoles in the brain over time (Figures 6A–6P; Figure S6S). Interestingly, *azot*^{−/−} brains showed higher number of vacuoles compared to control flies (*azot*^{+/+} and *azot*^{−/−}; *azot*^{+/+}) and a higher rate of vacuole accumulation developing over time (Figures 6N–6P). In the case of flies with three genomic copies of the gene (*azot*^{+/+}; *azot*⁺), vacuole number tended to be the lowest (Figures 6E, 6I, and 6M–6P).

Next, we analyzed the cumulative expression of *azot* during aging of the adult brain. We detected positive cells as revealed by reporter *azot*{*KO*; *gfp*}, in homozygosis, that express GFP instead of Azot. We observed a time-dependent accumulation of *azot*-positive cells (Figures 6Q–6W).

From this, we conclude that *azot* is required to prevent tissue degeneration in the adult brain and lack of *azot* showed signs of accelerated aging. This suggested that *azot* could affect the longevity of adult flies (Figures 6X and 6Y). We found that flies

lacking *azot* (*azot*^{−/−}) had a shortened lifespan with a median survival of 7.8 days, which represented a 52% decrease when compared to WT flies (*azot*^{+/+}), and a maximum lifespan of 18 days, 25% less than WT flies (*azot*^{+/+}). This effect on lifespan was *azot* dependent because it was completely rescued by introducing two functional copies of *azot* (Figures 6X and 6Y). On the contrary, flies with three functional copies of the gene (*azot*^{+/+}; *azot*⁺) showed an increase in median survival and maximum lifespan of 54% and 17%, respectively.

In conclusion, *azot* is necessary and sufficient to slow down aging, and active selection of viable cells is critical for a long lifespan in multicellular animals.

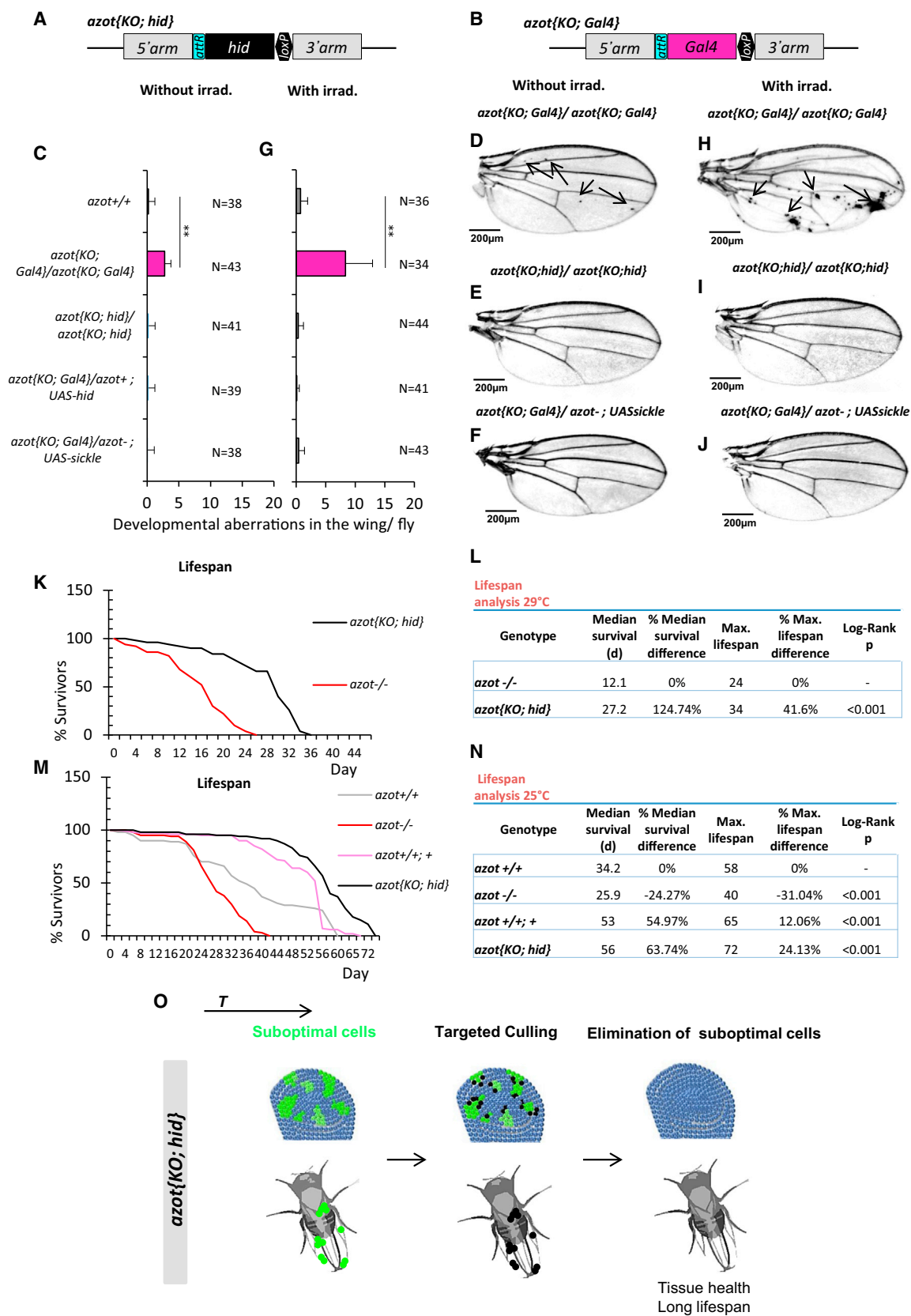
Death of Unfit Cells Is Sufficient and Required for Multicellular Fitness Maintenance

Our results show the genetic mechanism through which cell selection mediates elimination of suboptimal but viable cells. However, using flip-out clones and MARCM (Lee and Luo, 2001), we found that Azot overexpression was not sufficient to induce cell death in wing imaginal discs (Figures S6T–S6Y). Because Hid is downstream of Azot, we wondered whether expressing Hid under the control of the *azot* regulatory regions could substitute for Azot function.

In order to test this hypothesis, we replaced the whole endogenous *azot* protein-coding sequence by the cDNA of the proapoptotic gene *hid* (*azot*{*KO*; *hid*} flies; see Figure 7A). In a second strategy, the whole endogenous *azot* protein-coding sequence was replaced by the cDNA of transcription factor *Gal4*, so that the *azot* promoter can activate any *UAS* driven transgene (*azot*{*KO*; *Gal4*} flies (Figure 7B). We then compared the number of morphological aberrations in the adult wings of six genotypes: first, homozygous *azot*{*KO*; *Gal4*} flies that lacked Azot; second, *azot*{*KO*; *hid*} homozygous flies that express Hid with the *azot* pattern in complete absence of Azot; third, *azot*^{+/+} WT flies as a control; and finally three genotypes where the *azot*{*KO*; *Gal4*} flies were crossed with *UAS**hid*, *UAS**sickle*, another proapoptotic gene (Srinivasula et al., 2002), or *UAS**p35*, an apoptosis inhibitor. In the case of *UAS**sickle* flies, we introduced a second *azot* mutation to eliminate *azot* function. Interestingly, the number of morphological aberrations was brought back to WT levels in all the situations where the *azot* promoter was driving proapoptotic genes (*azot*{*KO*; *hid*}, *azot*{*KO*; *Gal4*} × *UAS**hid*, *azot*{*KO*; *Gal4*} × *UAS**sickle*, see Figures 7A–7J) with or without irradiation. On the contrary, expressing *p35* with the *azot* promoter was sufficient to produce morphological aberrations despite the presence of one

Figure 6. *azot* Is Required to Prevent Tissue Degeneration in the Adult Brain and to Promote Lifespan

(A–P) Brain integrity studies over time. (A) Axial plane of *Drosophila* WT brain counterstained with toluidine blue. (B–M) Magnification images of the central brain, counterstained with toluidine blue, showing degenerative vacuoles (white dots) of the following four genotypes over time: (1) *azot*^{+/+}, (2) *azot*^{−/−}, (3) *azot*^{−/−}; *azot*^{+/+}, and (4) *azot*^{+/+}; *azot*⁺. (N–P) Number of neurodegenerative vacuoles. (N) Number of degenerative vacuoles per brain area (70 × 70 μm) after 1 day at 29°C (*azot*^{+/+} n = 14, *azot*^{−/−} [p < 0.01] n = 8, *azot*^{−/−}; *azot*^{+/+} n = 16 and *azot*^{+/+}; *azot*⁺ [p < 0.01] n = 11). (O) Number of degenerative vacuoles per brain area after 7 days at 29°C (*azot*^{+/+} n = 16, *azot*^{−/−} [p < 0.01] n = 16, *azot*^{−/−}; *azot*^{+/+} n = 7 and *azot*^{+/+}; *azot*⁺ [p < 0.01] n = 20). (P) Number of degenerative vacuoles per brain area after 14 days at 29°C (*azot*^{+/+} n = 7, *azot*^{−/−} [p < 0.01] n = 3, *azot*^{−/−}; *azot*^{+/+} n = 10 and *azot*^{+/+}; *azot*⁺ n = 7). (Q–V) Azot-positive cells (green, GFP) in *azot*{*KO*; *gfp*} homozygous flies after 1 day (Q and R), 7 days (S and T), and 14 days (U and V) at 29°C. DAPI is in blue. (W) Number of Azot-positive cells per brain area (50 × 50 μm) in *azot*{*KO*; *gfp*} homozygous flies after 1 day (n = 11), 7 days (n = 15), and 14 days (n = 18) at 29°C. (X) Lifespan studies of the same four genotypes at 29°C. (Y) Lifespan values, including median survival and maximum lifespan, for the four genotypes. Data are represented as mean ± SEM.



(legend on next page)

functional copy of *azot* (Figures S7A–S7H). Likewise, *p35*-expressing flies (*azot*{*KO*; *Gal4*}/*azot*⁺; *UASp35*) did not survive UV treatments (Figure S7I), whereas a percentage of the flies expressing *hid* (26%) or *sickle* (28%) in *azot*-positive cells were able to survive (Figure S7I).

From this, we conclude that specifically killing those cells selected by the *azot* promoter is sufficient and required to prevent morphological malformations and provide resistance to UV irradiation.

Death of Unfit Cells Extends Lifespan

Next, we checked if the shortened longevity observed in *azot*^{−/−} flies could be also rescued by killing *azot*-expressing cells with *hid* in the absence of Azot protein. We found that *azot*{*KO*; *hid*} homozygous flies had dramatically improved lifespan with a median survival of 27 days at 29°C, which represented a 125% increase when compared to *azot*^{−/−} flies, and a maximum lifespan of 34 days, 41% more than mutant flies (Figures 7K and 7L).

Similar results were obtained at 25°C (Figures 7M and 7N). We found that flies lacking *azot* (*azot*^{−/−}) had a shortened lifespan with a median survival of 25 days, which represented a 24% decrease when compared to WT flies (*azot*^{+/+}), and a maximum lifespan of 40 days, 31% less than WT flies (*azot*^{+/+}). On the contrary, flies with three functional copies of the gene (*azot*^{+/+}; *azot*⁺) or flies where *azot* is replaced by *hid* (*azot*{*KO*; *hid*} homozygous flies) showed an increase in median survival of 54% and 63% and maximum lifespan of 12% and 24%, respectively.

Finally, we tested the effects of dietary restriction on longevity of those flies (Partridge et al., 2005) (Figures S7J and S7K). We found that dietary restriction could extend both the median survival and the maximum lifespan of all genotypes (Figures S7J and S7K). Interestingly, dietary restricted flies with three copies of the gene *azot* showed a further increase in maximum lifespan of 35% (Figure S7K). This shows that dietary restriction and elimination of unfit cells can be combined to maximize lifespan.

In conclusion, eliminating unfit cells is sufficient to increase longevity, showing that cell selection is critical for a long lifespan in *Drosophila*.

DISCUSSION

Here, we show that active elimination of unfit cells is required to maintain tissue health during development and adulthood. We

identify a gene (*azot*), whose expression is confined to suboptimal or misspecified but morphologically normal and viable cells. When tissues become scattered with suboptimal cells, lack of *azot* increases morphological malformations and susceptibility to random mutations and accelerates age-dependent tissue degeneration. On the contrary, experimental stimulation of *azot* function is beneficial for tissue health and extends lifespan. Therefore, elimination of less fit cells fulfils the criteria for a hallmark of aging (López-Otín et al., 2013).

Although cancer and aging can both be considered consequences of cellular damage (Greaves and Maley, 2012; López-Otín et al., 2013), we did not find evidence for fitness-based cell selection having a role as a tumor suppressor in *Drosophila*. Our results rather support that accumulation of unfit cells affect organ integrity and that, once organ function falls below a critical threshold, the individual dies.

We find Azot expression in a wide range of “less fit” cells, such as WT cells challenged by the presence of “supercompetitors,” slow proliferating cells confronted with normal proliferating cells, cells with mutations in several signaling pathways (i.e., Wingless, JAK/STAT, Dpp), or photoreceptor neurons forming incomplete ommatidia. In order to be expressed specifically in “less fit” cells, the transcriptional regulation of *azot* integrates fitness information from at least three levels: (1) the cell’s own levels of *Flower*^{Lose} isoforms, (2) the levels of *Sparc*, and (3) the levels of *Lose* isoforms in neighboring cells. Therefore, Azot ON/OFF regulation acts as a cell-fitness checkpoint deciding which viable cells are eliminated. We propose that by implementing a cell-fitness checkpoint, multicellular communities became more robust and less sensitive to several mutations that create viable but potentially harmful cells. Moreover, *azot* is not involved in other types of apoptosis, suggesting a dedicated function, and—given the evolutionary conservation of Azot—pointing to the existence of central cell selection pathways in multicellular animals.

EXPERIMENTAL PROCEDURES

In Situ Hybridization

We followed the protocol described in Rhiner et al. (2010). Probe sequences are available upon request.

Drosophila Genetics

Stocks and crosses were kept at 25°C in standard media. The following stocks were used: *ywf;tub > dmyc > Gal4/Cyo*; *UASgfp*; *azot::dsRed/TM6B*; *GMR-Gal4*; *azot::dsRed/TM6B*; *ywf;tub > dmyc > Gal4,azot^{−/−}/Cyo*; *UASgfp*; *ywf;tub*

Figure 7. Culling Azot-Expressing Cells Is Sufficient and Required for Multicellular Fitness Maintenance

(A and B) Knockin (KI) schemes (A) *azot*{*KO*; *Gal4*} and (B) *azot*{*KO*; *hid*}.

(C–F) Wings from 10- to 13-day-old flies and quantification of developmental aberrations of the following five genotypes: (C) *azot*^{+/+}, (C and D) *azot*{*KO*; *Gal4*}/*azot*{*KO*; *Gal4*}, (C and E) *azot*{*KO*; *hid*}/*azot*{*KO*; *hid*}, (C and F) *azot*{*KO*; *Gal4*}/*azot*^{−/−}; *UASsickle*, and (C) *azot*{*KO*; *Gal4*}/*azot*⁺; *UAShid*.

(G–J) Wings from 10- to 13-day-old flies and quantification of developmental aberrations after UV irradiation of the same five genotypes. Irradiation dose of 2 × 10^{−2} J administered during pupal stage 0.

(K and L) Comparative lifespan studies of genotypes *azot*{*KO*; *hid*}/*azot*{*KO*; *hid*} and *azot*^{−/−} at 29°C.

(L) Median and maximum survival of genotypes *azot*{*KO*; *hid*}/*azot*{*KO*; *hid*} and *azot*^{−/−}.

(M and N) Lifespan studies at 25°C of the following four genotypes: (1) *azot*^{+/+}, (2) *azot*^{−/−}, (3) *azot*^{+/+}; *azot*⁺, and (4) *azot*{*KO*; *hid*}/*azot*{*KO*; *hid*}. (N) Median and maximum survival of the four genotypes.

(O) Scheme showing that specifically killing Azot-expressing cells with the general proapoptotic factor *Hid* is sufficient to prevent morphological malformations and rescue *azot* mutant phenotypes.

Data are represented as mean ± SEM.

> dmyc > Gal4, azot⁻/Cyo; UASrpf; ywf; act > y+ > Gal4, azot⁻/Cyo; UASgfp; ywf; act > y+ > Gal4/Cyo; UASRNAiazot; azot{KO; gfp}; azot{KO; hid}; azot{KO; Gal4}; UASbrk; act > cd2 > Gal4, UASgfp/TM6B; act > y+ > Gal4, UASgfp; azot::dsRed/TM6B; w; flowerUbi-YFP, flowerLose-A-GFP, flowerLose-B-RFP; ywf; Ubigfp, MinuteFRT42/Cyo; ywf; FRT42/Cyo; hsFlp, UAS-CD8-GFP; GAL80 FRT 40A/Cyo; tub > G4/TM6B; ywf; Flp; armZFRT40A/Cyo; MKRS/TM6B; ywf; patched-Gal4; apterous-Gal4; GMR-Gal4, UASeiger; RNAifwe^{Lose^h} (Merino et al., 2013); ywf; UASmfwe³; ywf; UASsparc/TM6B; UASfwe^{Lose-B}; UASfwe^{Lose-A}; UASfwe^{Lose-A}, UASfwe^{Lose-B}; UASp35; UASpuckered; UASdAxi/TM3; UAShid; UASsickle; UASbax; UASbcl2; UASp53DN; UASRNAifweGD; UASRNAis parc(16678); UASRNAiazotGD(18166); UASRNAiazotKK(102353); UASRNAis cribble(Bloomington); UASRNAidlg(Bloomington); UASRNAihopscotch(Bloomington); UASRNAieigerGD; ywf; Cyo/ff; UASazot/TM6B; ywf; Cyo/ff; UASazot-HA/TM6B; ywf; Cyo/ff; UASazotpm4Q12/TM6B; ywf; UASlacZ; and UASCSK-IR.

Clone Induction

Flip-out clones were generated after heat shock at 37°C between 5 and 15 min. For ubiquitous expression experiments larvae were subjected to 45 min heat shock for all cells to perform flip-out and activate Gal4 under the control of the *actin* promoter (*act>Gal4*).

Azot Reporter: azot::dsRed

The genomic region 3 kb upstream plus the full exon was cloned in pRe-dStinger vector using XbaI and KpnI restriction sites. Primer sequences are available upon request.

Overexpressing Constructs

cDNA of azot was fully sequenced and subcloned into the pUASattB vector using XbaI and KpnI restriction sites. In order to generate N- and C-terminal HA-tagged forms, the respective cDNAs were amplified with primers containing the HA sequence and subcloned into KpnI and XbaI sites of pUASattB. Primer sequences are available upon request.

Azotpm4Q12

Site-directed mutagenesis was used to create point mutations that changed glutamic acid (E) to glutamine (Q) as shown in Figure S1A. Primer sequences are available upon request.

Azot Knockout Generation

We followed the genomic engineering strategy described in Huang et al. (2009); homologous regions are shown in (Figure 1A). Primer sequences are available upon request.

Knockin Generation

Knockout founder line (Figure 2A) was used for the generation of knockin flies as described in Huang et al. (2009). cDNA of *gfp*, *hid*, and *Gal4* was used for the generation of *azot{KO; gfp}*, *azot{KO; hid}*, and *azot{KO; Gal4}* knockin lines. Primer sequences are available upon request.

Immunohistochemistry

Standard immunohistochemistry protocol was used for antibody detection (Rhiner et al., 2010). For the generation of specific antibodies against Azot, N-terminal peptide MEDISHEERVLLDTRF was used to immunize rabbits. Anti-Wingless (ms, 1:50) was from DSHB, anti-caspase-3 (rabbit, 1:100) was from Cell Signaling Technology, anti-KDEL (rabbit, 1:100) was from Abcam, anti-cytochrome c (mouse, 1:800) was from BD Pharmingen, anti-Hid (rabbit, 1:50) and anti-HA (rat, 1:250) were from Roche, and anti-βGal (mouse, 1:200) was from Promega. TUNEL staining performed as described (Lolo et al., 2012). Confocal images acquired with Leica SP2 and SP5 microscopes.

UV Treatments

Treatments were performed using a UV Stratalinker 2400 machine (UV-B 254 nm). Adult flies were subjected to 2×10^{-2} J dose of UV irradiation when they were 1–3 days old and analyzed for Azot and Flower isoform expression 24 hr later. For lifespan experiments after irradiation, a dose of 5×10^{-2} J was used. Larvae and pupae were subjected to 2×10^{-2} J dose of

UV irradiation, and Azot expression or developmental aberrations were analyzed.

Longevity Assays

Cohorts of 100 female flies (1–3 days old) of the same genetic background were collected and kept at 29°C or 25°C on standard food (3.4 l water, 280 g maize, 36 g agar, 120 g yeast, 300 g sugar syrup, 32 g potassium, 6 g methyl, 20 ml propionic acid). Surviving flies were counted every 2 days (He and Jasper, 2014).

Dietary Restriction Assays

Cohorts of 100 female flies (1–3 days old) were collected and kept at 29°C on water-diluted standard food (one to one). Surviving flies were counted every 2 days.

Brain Studies

Brain Integrity

Adult flies kept at 29°C of the selected time points and genotypes were analyzed for the appearance of neurodegenerative vacuoles over time in the central brain as previously described (Kretschmar et al., 1997).

Azot Expression

Adult flies *azot{KO; gfp}/azot{KO; gfp}* were kept at 29°C. The selected time points were analyzed for the number of GFP-positive cells in the central brain.

Statistical Analysis

For the rescue assay using azot KO in supercompetition (Figure 2E), rescue assay in supercompetition with azot RNAi and overexpression of the protein (Figures S2J–S2P), the rescue assay of clones with apicobasal defects and the clones with deficient Wg signaling (Figures S6N–S6R), and brain integrity studies over time (Figures 6A–6P), the data were analyzed with the K independent samples test. The post hoc DMS test was then used to detect significant differences.

For the caspase-positive cells in *azot^{+/+}* and *azot^{-/-}* background (Figure 2D), the rescue assay in overexpression of Flower^{Lose} isoforms (Figures 2R–2T; Figure S2T), and azot overexpression in clones (Figures S6T–S6Y), all data were analyzed with two independent samples test (Mann-Whitney U test). Levene test was used to analyze number of cleaved caspase-3-positive cells, rescue assay of Flower^{Lose} isoforms, and number of azot-overexpressing clones.

For the quantification of the number of developmental aberrations before and after irradiation treatment in *azot^{+/+}*, *azot^{+/-}*, and *azot^{-/-}*, and *azot^{-/-}*; *azot^{+/+}* background (Figures 3A–3E, 3L–3P, 7C–7J, and S7A–S7H), data were analyzed with the K independent samples test (Levene), and Levy-Tukey was used for post hoc analyses.

In the rescue assay in supercompetition using RNAi (24 hr ACI) (Figures S2C–S2I), the data were analyzed with ANOVA test.

In the quantification of eye size in apoptosis assay (Figures S3H–S3N), the data were analyzed with ANOVA. Bonferroni post hoc test was used to detect significant differences among genotypes.

For the functional assays of azot in retinas (Figures 2G–2L), azot dose sensitive (Figures 2U–2W), rescue assay in overexpression of mouse flower³ isoform (Figure S2U), and rescue assay of clones with apicobasal defects, and clones with deficient Wg signaling by azot RNAi (Figures S6D–S6M), all data were analyzed with Student's t test.

For the lifespan analysis (Figures 6X, 7K, 7M, and S7J), the log-rank test was used to study significant differences among the genotypes.

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.12.017>.

ACKNOWLEDGMENTS

We thank H. Bellen, H. Stocker, M. Vidal, H. Ryoo, A. Bergmann, B. Mignotte, M. Portela, VDR, and Hybridoma bank and Bloomington for flies and antibodies; CONSOLIDER consortium for injecting flies; and members of our laboratory for comments on the manuscript. Work in our laboratory is funded by

the European Research Council, Swiss National Science Foundation, Josef Steiner Cancer Research Foundation, the bilateral Japanese-Swiss ST, and Swiss Cancer League.

Received: June 20, 2014

Revised: September 26, 2014

Accepted: November 24, 2014

Published: January 15, 2015

REFERENCES

- Baillon, L., and Basler, K. (2014). Reflections on cell competition. *Semin. Cell Dev. Biol.* 32, 137–144.
- Brodsky, M.H., Nordstrom, W., Tsang, G., Kwan, E., Rubin, G.M., and Abrams, J.M. (2000). Drosophila p53 binds a damage response element at the reaper locus. *Cell* 101, 103–113.
- de Beco, S., Ziosi, M., and Johnston, L.A. (2012). New frontiers in cell competition. *Developmental dynamics* 241, 831–841.
- de la Cova, C., Abril, M., Bellosta, P., Gallant, P., and Johnston, L.A. (2004). Drosophila myc regulates organ size by inducing cell competition. *Cell* 117, 107–116.
- Fernández-Hernández, I., Rhiner, C., and Moreno, E. (2013). Adult neurogenesis in Drosophila. *Cell Rep.* 3, 1857–1865.
- Fuchs, Y., and Steller, H. (2011). Programmed cell death in animal development and disease. *Cell* 147, 742–758.
- Gaumer, S., Guénal, I., Brun, S., Théodore, L., and Mignotte, B. (2000). Bcl-2 and Bax mammalian regulators of apoptosis are functional in Drosophila. *Cell Death Differ.* 7, 804–814.
- Greaves, M., and Maley, C.C. (2012). Clonal evolution in cancer. *Nature* 487, 306–313.
- He, Y., and Jasper, H. (2014). Studying aging in Drosophila. *Methods* 68, 129–133.
- Hogan, C., Kajita, M., Lawrenson, K., and Fujita, Y. (2011). Interactions between normal and transformed epithelial cells: their contributions to tumorigenesis. *Int. J. Biochem. Cell Biol.* 43, 496–503.
- Huang, J., Zhou, W., Dong, W., Watson, A.M., and Hong, Y. (2009). From the cover: directed, efficient, and versatile modifications of the Drosophila genome by genomic engineering. *Proc. Natl. Acad. Sci. USA* 106, 8284–8289.
- Igaki, T., Pastor-Pareja, J.C., Aonuma, H., Miura, M., and Xu, T. (2009). Intrinsic tumor suppression and epithelial maintenance by endocytic activation of Eiger/TNF signaling in Drosophila. *Dev. Cell* 16, 458–465.
- Jacobs, K.B., Yeager, M., Zhou, W., Wacholder, S., Wang, Z., Rodriguez-Santiago, B., Hutchinson, A., Deng, X., Liu, C., Horner, M.J., et al. (2012). Detectable clonal mosaicism and its relationship to aging and cancer. *Nat. Genet.* 44, 651–658.
- Kennedy, S.R., Loeb, L.A., and Herr, A.J. (2012). Somatic mutations in aging, cancer and neurodegeneration. *Mech. Ageing Dev.* 133, 118–126.
- Kretschmar, D., Hasan, G., Sharma, S., Heisenberg, M., and Benzer, S. (1997). The swiss cheese mutant causes glial hyperwrapping and brain degeneration in Drosophila. *The Journal of neuroscience* 17, 7425–7432.
- Laurie, C.C., Laurie, C.A., Rice, K., Doheny, K.F., Zelnick, L.R., McHugh, C.P., Ling, H., Hetrick, K.N., Pugh, E.W., Amos, C., et al. (2012). Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat. Genet.* 44, 642–650.
- Lee, T., and Luo, L. (2001). Mosaic analysis with a repressible cell marker (MARCM) for Drosophila neural development. *Trends Neurosci.* 24, 251–254.
- Lemaître, B., and Miguel-Aliaga, I. (2013). The digestive tract of Drosophila melanogaster. *Annu. Rev. Genet.* 47, 377–404.
- Levi-Montalcini, R. (1987). The nerve growth factor 35 years later. *Science* 237, 1154–1162.
- Liu, N., Landreh, M., Cao, K., Abe, M., Hendriks, G.J., Kennerdell, J.R., Zhu, Y., Wang, L.S., and Bonini, N.M. (2012). The microRNA miR-34 modulates ageing and neurodegeneration in Drosophila. *Nature* 482, 519–523.
- Lolo, F.N., Casas-Tintó, S., and Moreno, E. (2012). Cell competition time line: winners kill losers, which are extruded and engulfed by hemocytes. *Cell Rep.* 2, 526–539.
- López-Otín, C., Blasco, M.A., Partridge, L., Serrano, M., and Kroemer, G. (2013). The hallmarks of aging. *Cell* 153, 1194–1217.
- Martin-Blanco, E., Gampel, A., Ring, J., Virdee, K., Kirov, N., Tolkovsky, A.M., and Martinez-Arias, A. (1998). puckered encodes a phosphatase that mediates a feedback loop regulating JNK activity during dorsal closure in Drosophila. *Genes Dev.* 12, 557–570.
- Maune, J.F., Klee, C.B., and Beckingham, K. (1992). Ca²⁺ binding and conformational change in two series of point mutations to the individual Ca(2+)-binding sites of calmodulin. *J. Biol. Chem.* 267, 5286–5295.
- Merino, M.M., Rhiner, C., Portela, M., and Moreno, E. (2013). “Fitness fingerprints” mediate physiological culling of unwanted neurons in Drosophila. *Curr. Biol.* 23, 1300–1309.
- Morata, G., and Ripoll, P. (1975). Minutes: mutants of Drosophila autonomously affecting cell division rate. *Dev. Biol.* 42, 211–221.
- Moreno, E. (2014). Cancer: Darwinian tumour suppression. *Nature* 509, 435–436.
- Moreno, E., and Basler, K. (2004). dMyc transforms cells into super-competitors. *Cell* 117, 117–129.
- Moreno, E., and Rhiner, C. (2014). Darwin’s multicellularity: from neurotrophic theories and cell competition to fitness fingerprints. *Curr. Opin. Cell Biol.* 31C, 16–22.
- Moreno, E., Basler, K., and Morata, G. (2002). Cells compete for decapentaplegic survival factor to prevent apoptosis in Drosophila wing development. *Nature* 416, 755–759.
- Moskalev, A.A., Shaposhnikov, M.V., Plyusnina, E.N., Zhavoronkov, A., Budovsky, A., Yanai, H., and Fraifeld, V.E. (2013). The role of DNA damage and repair in aging through the prism of Koch-like criteria. *Ageing Res. Rev.* 12, 661–684.
- Parisi, F., Stefanatos, R.K., Strathdee, K., Yu, Y., and Vidal, M. (2014). Transformed epithelia trigger non-tissue-autonomous tumor suppressor response by adipocytes via activation of Toll and Eiger/TNF signaling. *Cell Rep.* 6, 855–867.
- Partridge, L., Piper, M.D., and Mair, W. (2005). Dietary restriction in Drosophila. *Mech. Ageing Dev.* 126, 938–950.
- Petrova, E., López-Gay, J.M., Rhiner, C., and Moreno, E. (2012). Flower-deficient mice have reduced susceptibility to skin papilloma formation. *Dis. Model. Mech.* 5, 553–561.
- Portela, M., Casas-Tinto, S., Rhiner, C., López-Gay, J.M., Domínguez, O., Soldini, D., and Moreno, E. (2010). Drosophila SPARC is a self-protective signal expressed by loser cells during cell competition. *Dev. Cell* 19, 562–573.
- Raff, M.C. (1992). Social controls on cell survival and cell death. *Nature* 356, 397–400.
- Reeves, H.M. (2006). Sahagún’s “Florentine codex,” a little known Aztec natural history of the Valley of Mexico. *Arch. Nat. Hist.* 33, 302–321.
- Rhiner, C., López-Gay, J.M., Soldini, D., Casas-Tinto, S., Martín, F.A., Lombardía, L., and Moreno, E. (2010). Flower forms an extracellular code that reveals the fitness of a cell to its neighbors in Drosophila. *Dev. Cell* 18, 985–998.
- Rodrigues, A.B., Zoranovic, T., Ayala-Camargo, A., Grewal, S., Reyes-Robles, T., Krasny, M., Wu, D.C., Johnston, L.A., and Bach, E.A. (2012). Activated STAT regulates growth and induces competitive interactions independently of Myc, Yorkie, Wingless and ribosome biogenesis. *Development* 139, 4051–4061.
- Simi, A., and Ibáñez, C.F. (2010). Assembly and activation of neurotrophic factor receptor complexes. *Dev. Neurobiol.* 70, 323–331.
- Simpson, P. (1979). Parameters of cell competition in the compartments of the wing disc of Drosophila. *Dev. Biol.* 69, 182–193.
- Srinivasula, S.M., Datta, P., Kobayashi, M., Wu, J.W., Fujioka, M., Hegde, R., Zhang, Z., Mukattash, R., Fernandes-Alnemri, T., Shi, Y., et al. (2002). sickle,

a novel *Drosophila* death gene in the reaper/hid/grim region, encodes an IAP-inhibitory protein. *Curr. Biol.* **12**, 125–130.

Suzanne, M., Petzoldt, A.G., Speder, P., Coutelis, J.B., Steller, H., and Noselli, S. (2010). Coupling of apoptosis and L/R patterning controls stepwise organ looping. *Curr. Biol.* **20**, 1773–1778.

Szilard, L. (1959). On the Nature of the Aging Process. *Proc. Natl. Acad. Sci. USA* **45**, 30–45.

Tamori, Y., and Deng, W.M. (2011). Cell competition and its implications for development and cancer. *Journal of genetics and genomics* **38**, 483–495.

Tamori, Y., Bialucha, C.U., Tian, A.G., Kajita, M., Huang, Y.C., Norman, M., Harrison, N., Poulton, J., Ivanovitch, K., Disch, L., et al. (2010). Involvement of Lgl and Mahjong/VprBP in cell competition. *PLoS Biol.* **8**, e1000422.

Vanneste, E., Voet, T., Le Caignec, C., Ampe, M., Konings, P., Melotte, C., Debrock, S., Amyere, M., Vikkula, M., Schuit, F., et al. (2009). Chromosome instability is common in human cleavage-stage embryos. *Nat. Med.* **15**, 577–583.

Vidal, M., Larson, D.E., and Cagan, R.L. (2006). Csk-deficient boundary cells are eliminated from normal *Drosophila* epithelia by exclusion, migration, and apoptosis. *Dev. Cell* **10**, 33–44.

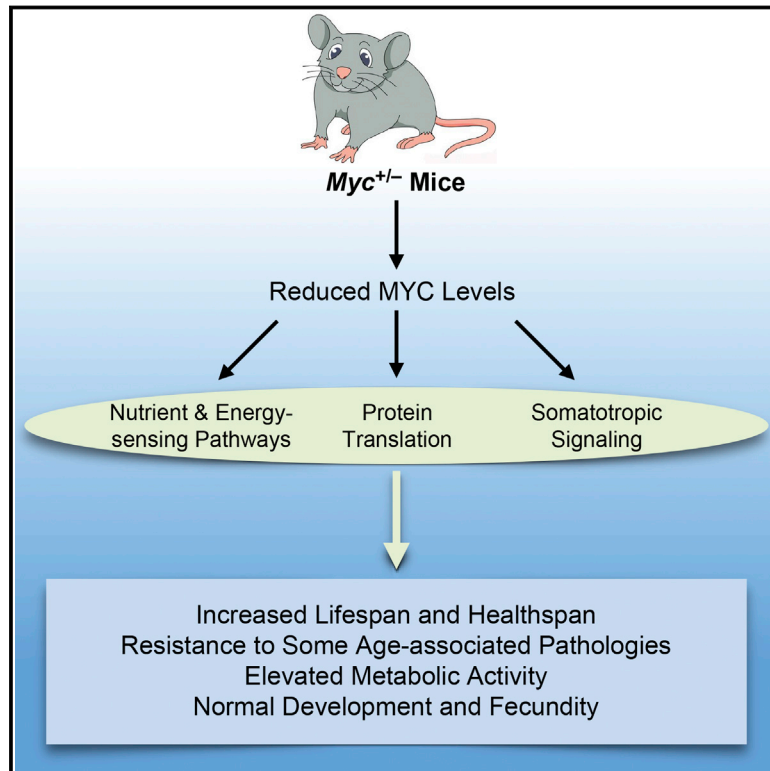
Vincent, J.P., Kolahgar, G., Gagliardi, M., and Piddini, E. (2011). Steep differences in wingless signaling trigger Myc-independent competitive cell interactions. *Dev. Cell* **21**, 366–374.

Willert, K., Logan, C.Y., Arora, A., Fish, M., and Nusse, R. (1999). A *Drosophila* Axin homolog, Daxin, inhibits Wnt signaling. *Development* **126**, 4165–4173.

Wolff, T., and Ready, D.F. (1991). Cell death in normal and rough eye mutants of *Drosophila*. *Development* **113**, 825–839.

Reduced Expression of MYC Increases Longevity and Enhances Healthspan

Graphical Abstract



Authors

Jeffrey W. Hofmann, Xiaoi Zhao, ..., Nicola Neretti, John M. Sedivy

Correspondence

john_sedivy@brown.edu

In Brief

Reduced expression of MYC increases lifespan in mice and benefits multiple aspects related to the aging process without apparent developmental trade-offs or changes in stress management pathways.

Highlights

- Reduction of *Myc* expression in mice promotes longevity
- Lower *Myc* levels benefit multiple organs and physiological processes
- Decreased *Myc* activity triggers changes in core nutrient and energy-sensing pathways
- Lower *Myc* levels do not cause apparent changes in stress management pathways

Accession Numbers

GSE55272



Reduced Expression of MYC Increases Longevity and Enhances Healthspan

Jeffrey W. Hofmann,^{1,7} Xiaoi Zhao,^{1,7} Marco De Cecco,¹ Abigail L. Peterson,¹ Luca Pagliaroli,¹ Jayameenakshi Manivannan,¹ Gene B. Hubbard,² Yuji Ikeno,² Yongqing Zhang,³ Bin Feng,⁴ Xiayi Li,⁵ Thomas Serre,⁵ Wenbo Qi,² Holly Van Remmen,² Richard A. Miller,⁶ Kevin G. Bath,⁵ Rafael de Cabo,³ Haiyan Xu,⁴ Nicola Neretti,¹ and John M. Sedivy^{1,*}

¹Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, RI 02912, USA

²Department of Cellular and Structural Biology, Barshop Institute for Longevity and Aging Studies, University of Texas Health Science Center at San Antonio, San Antonio, TX 78229, USA

³Translational Gerontology Branch, National Institute on Aging, 251 Bayview Boulevard, Suite 100, Baltimore, MD 21224, USA

⁴Hallett Center for Diabetes and Endocrinology, Rhode Island Hospital, Warren Alpert Medical School of Brown University, Providence, RI 02903, USA

⁵Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI 02912, USA

⁶Department of Pathology and Geriatrics Center, University of Michigan, Ann Arbor, MI 48109, USA

⁷Co-first author

*Correspondence: john_sedivy@brown.edu

<http://dx.doi.org/10.1016/j.cell.2014.12.016>

SUMMARY

MYC is a highly pleiotropic transcription factor whose deregulation promotes cancer. In contrast, we find that *Myc* haploinsufficient (*Myc*^{+/-}) mice exhibit increased lifespan. They show resistance to several age-associated pathologies, including osteoporosis, cardiac fibrosis, and immunosenescence. They also appear to be more active, with a higher metabolic rate and healthier lipid metabolism. Transcriptomic analysis reveals a gene expression signature enriched for metabolic and immune processes. The ancestral role of MYC as a regulator of ribosome biogenesis is reflected in reduced protein translation, which is inversely correlated with longevity. We also observe changes in nutrient and energy sensing pathways, including reduced serum IGF-1, increased AMPK activity, and decreased AKT, TOR, and S6K activities. In contrast to observations in other longevity models, *Myc*^{+/-} mice do not show improvements in stress management pathways. Our findings indicate that MYC activity has a significant impact on longevity and multiple aspects of mammalian healthspan.

INTRODUCTION

Myc is a helix-loop-helix leucine zipper transcription factor that is highly conserved among metazoans (Meyer and Penn, 2008). It was discovered as the transforming oncogene of the MC29 avian myelocytomatosis virus and subsequently as the cellular proto-oncogene activated in Burkitt's lymphoma. Increased expression of the MYC protein strongly promotes cell proliferation and has been documented as a frequent event in a wide variety of human cancers (Dang, 2012).

By interacting with partners such as MAX and ZBTB17 (MIZ1), MYC can either activate or repress transcription (Meyer and Penn, 2008). Much effort has been focused on understanding how MYC influences signaling networks and it has emerged as a major regulatory hub. In addition to its role in cancer, it is also critically involved with many essential cellular processes, and the mouse knockout is embryonic lethal. By conservative estimates, 15%–20% of all genes are directly regulated by MYC, including genes that play key roles in metabolism, ribosome biogenesis, cell cycle, apoptosis, differentiation, and stem cell maintenance (Dang, 2012).

While age does not have a significant effect on *Myc* expression in any mouse tissue examined (Zahn et al., 2007), many of the biological processes regulated by MYC have also been implicated in aging and age-associated diseases. MYC upregulates major biosynthetic pathways leading to cellular growth and proliferation and enhances energy production through glycolysis and oxidative phosphorylation (Dang, 2012). In contrast, calorie restriction (CR) and reduction of insulin/IGF-1 signaling promote longevity (Gems and Partridge, 2013). MYC also increases protein synthesis by positively regulating ribosome biogenesis (Brown et al., 2008), while reducing translation can extend lifespan (Johnson et al., 2013).

MYC overexpression results in an increase in reactive oxygen species (ROS) and DNA damage (Vafa et al., 2002), which are believed to contribute to the progression of aging (Hoeijmakers, 2009). Stem cell populations decline in number and functionality with normal aging (Cho et al., 2008; Jang et al., 2011), and ectopic MYC expression depletes stem cell populations (Eilers and Eisenman, 2008). MYC may also affect the inflammatory state that accompanies aging, because it directly regulates expression of some cytokines (Whitfield and Soucek, 2012) and may influence the composition of the leukocyte population via its roles in proliferation and stem cell maintenance (Eilers and Eisenman, 2008; Wang et al., 2011a).

The overall trend suggested by this evidence is that increased MYC activity promotes several processes that have been

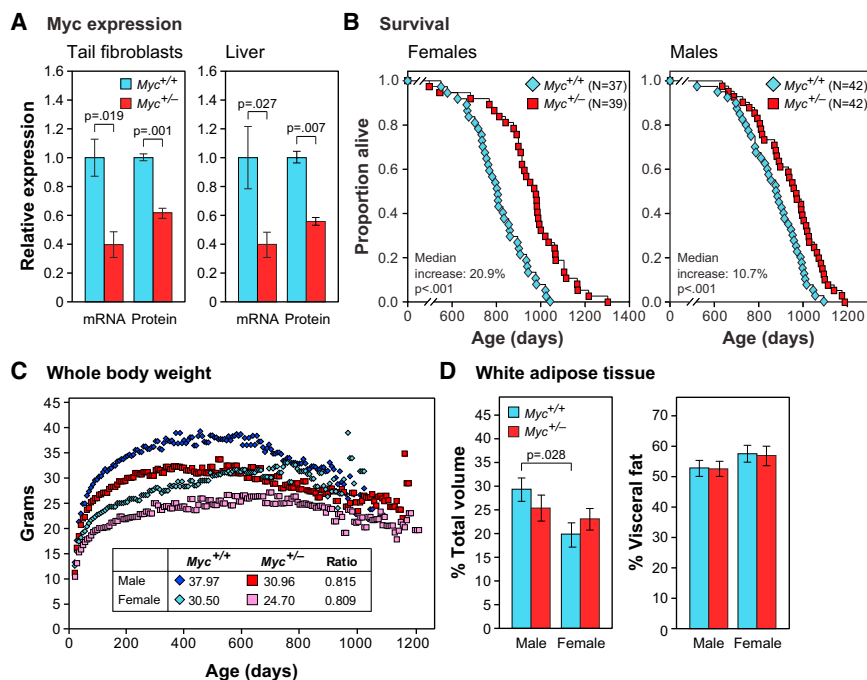


Figure 1. MYC Expression, Longevity, Body Mass, and Composition

(A) MYC mRNA and protein expression. mRNA was measured by qRT-PCR. Protein was measured by immunoblotting of tail fibroblasts or by IF of liver sections. Fibroblasts: n = 3, 3 months, females. Liver: n = 3–5, 5–9 months, both sexes.

(B) Survival of *Myc*^{+/+} (blue) and *Myc*^{+/-} (red) mice. Each data point shows one animal (N, number of animals in each cohort).

(C) Lifelong trends of whole body weights. Median weights (3 week sliding window) of the aging cohorts in (B). Inset: weights at 500 days and relative weights of *Myc*^{+/-} animals.

(D) Relative adiposity. Animals were scanned by micro-CT imaging. Left: volume of all white fat as percent of total volume of the animal. Right: volume of visceral fat as percent of total fat. n = 6, 5 months.

Error bars represent SEM.

See also Figure S1 and Tables S1 and S2.

connected with aging and age-associated diseases. To address the role of *Myc* in aging, given that a complete loss of *Myc* is embryonic lethal while overexpression promotes cancer, we established a partial loss of function (hypomorphic) model in the mouse. We previously found that cells knocked out for one copy of the *Myc* gene display a variety of mild but distinct phenotypes, including reduced rates of proliferation (Mateyak et al., 1997). Similarly, heterozygous (*Myc*^{+/-}) mice appear normal and healthy, but have a 20% smaller body mass (de Alboran et al., 2001; Trumpf et al., 2001). We used this simple constitutive hypomorphic model to address the consequences of reduced MYC activity on aging.

RESULTS

Experimental Strategy

Mice with all coding exons of one copy of the *Myc* gene flanked by LoxP sites (de Alboran et al., 2001) were bred to mice expressing germline Cre recombinase, converting the floxed allele to a deletion, and subsequently backcrossed to C57BL/6 for ten generations. The expected decreases of MYC mRNA and protein levels in *Myc*^{+/-} mice were confirmed in tail fibroblasts and several tissues (Figure 1A and Figures S1A–S1C available online). MYC chromatin immunoprecipitation of liver extracts from *Myc*^{+/-} animals reproducibly gave lower yields, indicating that in vivo a smaller fraction of DNA was bound by MYC (Figure S1D).

Myc^{+/-} Mice Have Increased Longevity

Large cohorts of both sexes and genotypes were maintained in a barrier facility and allowed to die of natural causes. A highly significant increase in median lifespan was observed: 10.7% for

males, 20.9% for females, and 15.1% for both sexes combined (Figure 1B; Tables S1 and S2). We do not know the reason for the greater effect in females; we note, however, that control (*Myc*^{+/+}) females were shorter lived than males (Figure S1E), an observation that has been reported in several colonies of C57BL/6 mice (Ladiges et al., 2009). Thus, while *Myc*^{+/+} males lived 8.8% longer than *Myc*^{+/-} females, *Myc*^{+/-} males and females had equivalent lifespans. Maximum lifespans were commensurately increased, with nearly all of the mice surviving to the longest-lived decile being of *Myc*^{+/-} genotype. The instantaneous mortality rate was lower for *Myc*^{+/-} mice across all ages (Figure S1F), indicating that the health benefits are not limited to a particular age.

Myc^{+/-} Mice Are Smaller and Develop and Reproduce Normally

Myc^{+/-} mice are healthy, robust, and can be group-housed with their *Myc*^{+/+} littermates without any adverse consequences. Animals of both sexes are 15%–20% smaller as adults (Figure 1C). These differences were apparent at weaning and the weight curves of *Myc*^{+/+} and *Myc*^{+/-} mice were parallel over time. Mass has been noted as a predictor of longevity in mice, although this effect is strain-specific (Anisimov et al., 2004). We found no correlation between longevity and the weight attained by individual mice within any cohort (Figure S1G).

The decreased size of *Myc*^{+/-} mice is proportional across all parts of the body, with the mass ratio of major organs to total body weight not significantly different between genotypes (Figure S1H). Prior examination of *Myc* hypomorphic mice did not find significant changes in cell size among several organs, which we confirmed in our animals (Figure S1I). Of particular interest was adipose tissue, because in several long-lived mouse models reduced adiposity has been associated with longevity. *Myc*^{+/-} and *Myc*^{+/+} mice have a similar proportion of adipose tissue relative to total body volume and also have similar proportions of

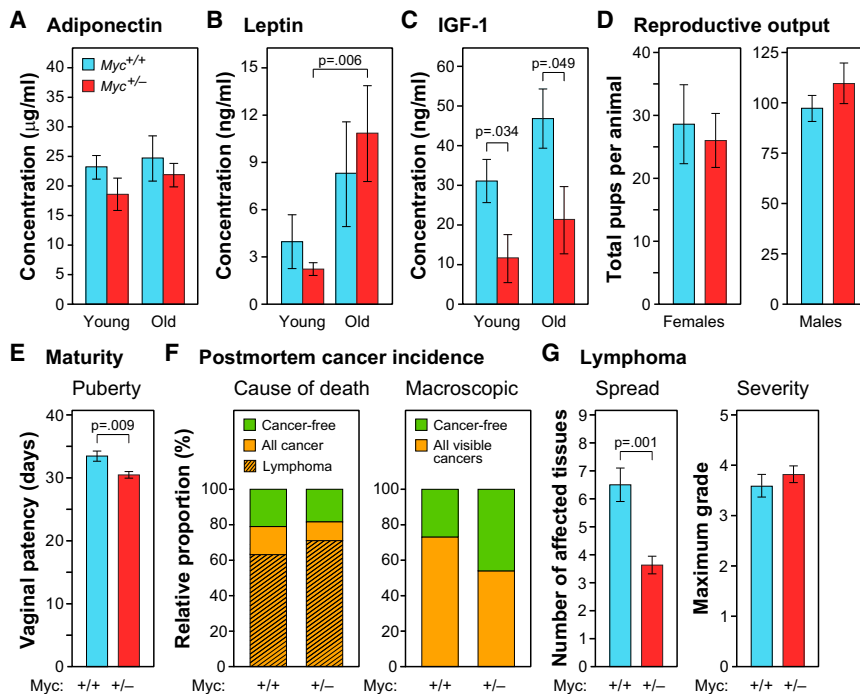


Figure 2. Metabolic Hormones, Fecundity, and Cancer Incidence

(A–C) Levels of adiponectin, leptin, and free IGF-1 in plasma. Blood samples were collected after overnight fasting. *n* = 5–12, 5 and 22 months, females.

(D) Lifetime reproductive output. *Myc*^{+/-} and *Myc*^{+/+} females (*n* = 5) were bred with *Myc*^{+/+} males (left), and *Myc*^{+/-} and *Myc*^{+/+} males (*n* = 7–9) were bred with *Myc*^{+/+} females (right).

(E) Sexual maturation of females. *n* = 12–19.

(F) Left: cause of death determined by a veterinary pathologist. *n* = 19 *Myc*^{+/+}, 38 *Myc*^{+/-}, both sexes. Right: incidence of macroscopic cancer noted at time of autopsy (*Myc*^{+/+}, 73.4%; *Myc*^{+/-}, 53.7%; *p* = 0.03). *n* = 64 *Myc*^{+/+}, 67 *Myc*^{+/-}, both sexes.

(G) Histopathological analysis was used to determine the spread (left; number of affected tissues) and severity (right; maximum grade) of lymphoma. *n* = 12 *Myc*^{+/+}, 27 *Myc*^{+/-}, both sexes.

Error bars represent SEM.

See also Figure S2 and Table S3.

visceral and subcutaneous adipose depots (Figure 1D). Consistent with this, *Myc*^{+/-} and *Myc*^{+/+} mice had similar levels of both leptin and adiponectin in serum at both 6 and 22 months, implying similar adipose tissue function (Figures 2A and 2B). The reduced size of *Myc*^{+/-} mice may be due to their lower levels of serum IGF-1, which we documented in both young and old animals (Figure 2C).

The concept of a tradeoff between longevity and reproduction, namely that increased longevity may bear the cost of decreased efficacy or duration of reproductive ability, has been widely discussed. We tested both sexes for reproductive fitness and longevity and found no significant differences between the two genotypes (Figures 2D and S2A–S2C). *Myc*^{+/-} females displayed slightly accelerated vaginal patency (Figure 2E), indicating that sexual maturation is not retarded. The increased longevity of *Myc*^{+/-} mice is thus not associated with either slower development or reduced fecundity. We also performed skin wound-healing assays and did not find statistically significant differences between the genotypes, even in old animals (Figure S2D).

The Major Cause of Death Is Lymphoma in Both Genotypes

We performed autopsies on all adequately preserved mice at time of death and a subset of animals were subjected to a histopathological analysis of 18 tissues. Seventy-nine percent of *Myc*^{+/+} and 81% of *Myc*^{+/-} mice were deemed to have died of cancers (of which 80% and 87%, respectively, were lymphomas; Figure 2F). The spectrum of pathologies was typical of the C57BL/6 strain of mice and very similar in both genotypes (Table S3). Hence, the increased longevity of *Myc*^{+/-} mice is not due to fewer cancer-caused deaths.

We, however, also noted evidence for decreased progression of cancer in *Myc*^{+/-} animals. First, careful macroscopic examination at time of autopsy revealed fewer tumors visible to the naked eye (Figure 2F). Second, histopathological analysis showed that by the time of death lymphoma spread to significantly fewer organs (Figure 2G). Third, while the severity (maximum grade) of lymphoma at time of death was similar in both genotypes, *Myc*^{+/-} mice survived considerably longer. Others have reported that reducing cancer prevalence alone does not substantially extend lifespan (Matheu et al., 2004). Additional data from multiple organ systems presented below indicate that other effects of the *Myc*^{+/-} genotype are likely to contribute to the extended lifespan.

Gene Expression Analysis Points to Changes in Metabolism, the Immune System, and a Unique Expression Signature

Microarray expression profiling was performed on liver, skeletal muscle, and white adipose (gonadal) tissues in 5- and 24-month-old animals. A principal component analysis showed that the first component was tissue, the second age, and the third genotype, representing 61.6%, 30.0%, and 2.9% of the total variability, respectively (Figure 3A; Table S4). The relatively small effect of the *Myc*^{+/-} genotype is notable, accounting for 10-fold fewer changes than those due to age. For all three tissues, both the number of differentially expressed genes and magnitude of average change with age were lower in *Myc*^{+/-} animals (Figure 3B). This reduction in the apparent aging of the transcriptome suggests that *Myc*^{+/-} mice are long-lived because of widespread changes that affect multiple tissues.

Differentially expressed genes were enriched in pathways related to the immune system and metabolism, especially that of lipids (Figure S3A). We also evaluated which upstream regulators could explain this pattern of transcription (Table S5). When sorted by significance for the effect of genotype in old animals,

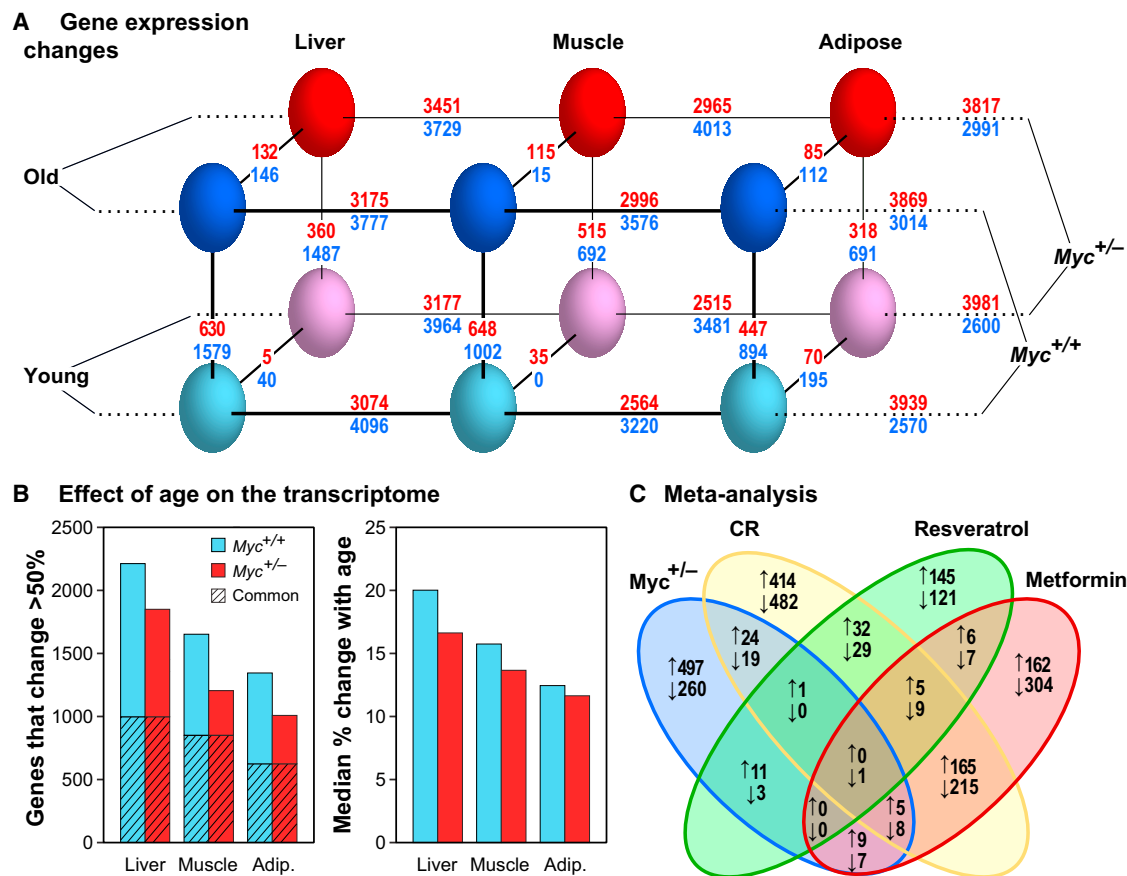


Figure 3. Transcriptome Analysis

(A) Gene expression changes. Three parameters were compared: genotype ($Myc^{+/+}$, $Myc^{+/-}$), age (5, 24 months), and tissue (liver, muscle, adipose). Genotype, age, and tissue comparisons are connected with diagonal, vertical, and horizontal lines, respectively. The number of genes changing expression is shown above and below each line. Red numbers: genes upregulated in $Myc^{+/+}$ versus $Myc^{+/-}$ animals (MYC-activated genes); blue numbers: genes upregulated in $Myc^{+/-}$ versus $Myc^{+/+}$ animals (MYC-repressed genes). A 1.5-fold cutoff and a FDR threshold of <5% were used. $n = 5$ –8 male animals.

(B) Effect of age on the transcriptome. Left: number of genes whose expression changes with age by more than 50% in either direction; hatched area represents genes in common between the two genotypes. Right: the median change in expression with age (expressed as %) across all expressed genes.

(C) Meta-analysis. Differentially expressed genes in $Myc^{+/-}$ versus $Myc^{+/+}$ animals were compared with genes similarly recovered in studies of calorie restriction, metformin, or resveratrol treatment (Martin-Montalvo et al., 2013; Pearson et al., 2008). Upward arrows indicate genes upregulated in the long-lived condition (and conversely for downward arrows). Expression in liver of old male mice was compared. Skeletal muscle showed the same trends. See also Figure S3 and Tables S4, S5, and S6.

regulators of lipid metabolism and the immune system were prominently enriched. Transcription factors that promote lipid biosynthesis and adipogenic fate were predicted to become more active with age in tissues in which adipogenesis and lipid accumulation are considered pathogenic, such as the liver and skeletal muscle, and less active with age in adipose tissue. These changes were reversed in old $Myc^{+/-}$ tissues. The effects of numerous immune system regulators (including interferons, interleukins, colony stimulating factors and NF- κ B) were strongly predicted to increase with age in all three tissues, and this effect was predicted to be counteracted in muscle of $Myc^{+/-}$ mice.

Expression of xenobiotic metabolism enzyme (XME) genes increases with age and is further elevated in male mice that are long-lived due to genetic, pharmacologic, or dietary interventions (female mice express these genes at an elevated and mostly constant level) (Li et al., 2013). The XME gene expression

signature in male $Myc^{+/-}$ mice did not resemble the signatures observed in several mouse longevity models (Table S6). This analysis is consistent with the results of stress tests performed in cell culture with tail fibroblasts established from adult $Myc^{+/-}$ mice (Figure S3D), which did not display increased resistance to any of the stressors previously shown to be counteracted by cells from a variety of long-lived mice. We also note that compared to other longevity models, some of which display very pronounced changes in XME gene expression, the changes in our animals were very modest (Table S6).

In a broader meta-analysis, we compared our gene expression data sets with those of other lifespan and healthspan extending interventions: CR and treatments with resveratrol or metformin. While some of these interventions show considerable overlap with one another in differentially expressed genes, especially CR and metformin (Martin-Montalvo et al., 2013), none were

similar to the *Myc*^{+/-} signature (Figures 3C and S3B). A meta-analysis based on Gene Ontology (GO) pathways also failed to indicate substantial overlap (Figure S3C). As before, CR and metformin showed much greater similarity to each other than either with *Myc*^{+/-}. The few universally enriched GO terms, however, further implicated metabolism and inflammation; for example, in muscle the only commonly enriched GO term was immune response, while in liver high density lipoprotein binding and hormone activity were two of only five enriched GO terms.

***Myc*^{+/-} Mice Show Evidence of Improved Healthspan in Multiple Tissues**

The pathological effects of aging in several organ systems were found to be attenuated in *Myc*^{+/-} mice. Cardiac fibrosis increases with age and in diabetes and obesity and has been found to be reduced in several long-lived mouse models (Dai et al., 2009). *Myc*^{+/-} mice had less cardiac fibrosis in old age than *Myc*^{+/+} animals (Figure 4A). An important component of healthspan in females is osteoporosis (Syed and Melim, 2011). In *Myc*^{+/+} females, both bone volume and trabecular number declined with age, whereas trabecular spacing increased (Figures 4B and 4C). Using all these parameters, old *Myc*^{+/-} mice were indistinguishable from young *Myc*^{+/+} animals, indicating that *Myc*^{+/-} females do not develop osteoporosis by 22 months of age. Musculoskeletal and neurological performance also decline with age, and the rotarod test is a frequently used measure of motor coordination that reflects these parameters. Old *Myc*^{+/-} mice were able to remain on the rotarod nearly twice as long as *Myc*^{+/+} animals of the same age (Figure 4G), indicating an attenuation of the effects of aging on motor function.

Lipid metabolism is strongly affected by aging. Hepatic lipid droplets (LDs) are small membrane-bound organelles that store fatty acids and cholesterol and regulate triglyceride metabolism (Fujimoto and Parton, 2011). LD activities occur mainly at their surface and hence their size rather than aggregate volume is a relevant measure of functional capacity. We found that while total hepatic LD area decreased ~2-fold in *Myc*^{+/+} animals with age (Figure S4A), average LD size increased by over 3.5-fold (Figure 4D). In contrast, total LD content in *Myc*^{+/-} mice did not change, and LD size increased to a lesser extent. The relative preservation of LD surface area in *Myc*^{+/-} animals is also consistent with increased expression of genes involved in their biogenesis (*Plin2*, *Cidec*, *Fitm1*).

Another important healthspan parameter of lipid metabolism is cholesterol. Cholesterol levels in mouse liver increase with age, and interventions that promote longevity, such as CR, result in decreased cholesterol synthesis (Tsuchiya et al., 2004). In our microarray data sets, the expression of cholesterol biosynthetic genes increased with age in *Myc*^{+/+} mice. In contrast, almost all genes in this pathway showed lower expression in *Myc*^{+/-} animals, which we confirmed by qRT-PCR (Figure 4E; this included the rate limiting enzyme, *Hmgcr*, and the master transcriptional regulator, *Srebf2*). Furthermore, both esterified and nonesterified cholesterol was reduced in serum of *Myc*^{+/-} mice, as well as in liver, where it is synthesized and stored (Figure 4F).

Immunosenescence results in part from thymic involution as well as replicative senescence of some T cell populations (Akbar and Henson, 2011). Because few T cells are produced by the

thymus in older animals, the T cell pool is maintained by the expansion of existing memory T cells, which is greater for cytotoxic CD8+ than regulatory CD4+ T cells (Goronzy and Weyand, 2005). Naive T cells do not proliferate. As previously reported, the ratios of both CD4+ to CD8+ T cells and of naive to memory T cells declined with age in normal (*Myc*^{+/+}) mice, but both were significantly elevated in *Myc*^{+/-} animals at 24 months of age (Figures 5B and 5C). We found no effect of genotype on total T cell levels (Figure 5A), the presence of macrophages in several tissues (Figure S4B), or total white blood cell counts (Figure S4C). These results indicate that *Myc*^{+/-} animals can maintain younger and presumably healthier proportions of T cell populations.

We next examined age-related changes in the bone marrow and thymus where T cells develop and mature. In C57BL/6 mice the thymus begins to involute at 10 months of age and very little remains by 20 months. Thymic mass was slightly decreased in *Myc*^{+/-} mice although the rate of involution was not changed; adiposity and the proportion of senescent cells were also not affected by genotype (Figures S4D and S4E). Hence, the younger T cell profiles in *Myc*^{+/-} animals are unlikely to be caused by reduced thymic involution.

Hematopoietic stem cells (HSC) differentiate through several subpopulations: long term HSC (LT-HSC) differentiate into short term HSC (ST-HSC), which can then differentiate into either common myeloid progenitors (CMP) or common lymphoid progenitors (CLP). With normal aging, the CLP/CMP ratio decreases by ~3-fold (Cho et al., 2008). Flow cytometry of HSC showed that 16-month-old *Myc*^{+/-} mice had a higher CLP/CMP ratio relative to *Myc*^{+/+} animals (Figures 5D and 5E). This is further evidence for decelerated aging of the hematopoietic system and might be responsible for the observed changes in T cell proportions. Consistent with this, we found that *Myc*^{+/-} mice have a lower ratio of ST-HSC to LT-HSC (Figure 5F), suggesting that LT-HSC can persist in greater numbers and/or functional capacity into old age.

***Myc*^{+/-} Mice Do Not Show Changes in Pathways Linked to Macromolecular Damage**

A commonly invoked cause of aging is the accumulation of damage to macromolecules, particularly from oxidative stress. Genotoxic stress, manifested as various forms of DNA damage, increases dramatically with age (Hoeijmakers, 2009). We examined the frequency of DNA damage foci, and while we saw clear evidence of the age-associated increase, the *Myc* genotype had no effect (Figure 5G). Genotoxic stress and other forms of damage can lead to apoptosis, which has been found to rise with age in several tissues (Kujoth et al., 2005). While we observed the increase with age, the changes were equivalent in *Myc*^{+/+} and *Myc*^{+/-} mice (Figure 5H). Genotoxic stress is also a major trigger of cellular senescence. As with apoptosis, however, *Myc* genotype did not affect the age-associated increase in cellular senescence (Figure S4E).

The cyclin-dependent kinase inhibitors p21 (*Cdkn1a*) and p16 (*Cdkn2a*) are important regulators of cellular senescence. As expected, both p21 and p16 were upregulated with aging (Figures S4F and S4G). The expression of p21, which responds strongly to DNA damage and ROS, was unaffected by *Myc* genotype in five tissues. The response of p16, whose regulation is not well

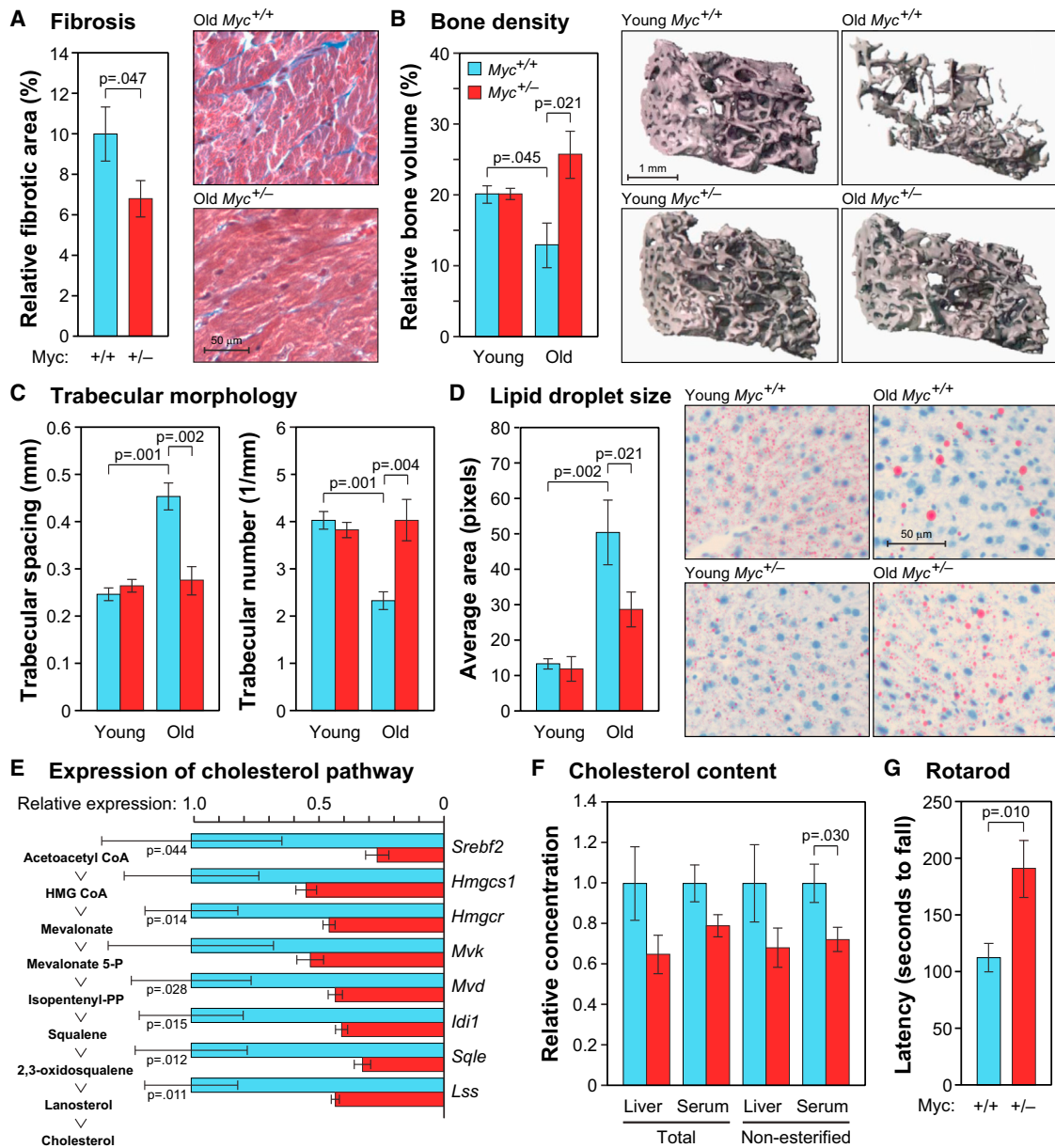


Figure 4. Amelioration of Age-Associated Phenotypes

(A) Cardiac fibrosis was scored in ventricular cross sections using Masson's trichrome stain. *n* = 11–14, 22–24 months, both sexes.

(B) Osteoporosis in females was assessed using micro-CT analysis. *n* = 3–7, 5 and 22 months.

(C) Trabecular spacing and number were scored by micro-CT, as above.

(D) Liver sections were stained with Oil Red O. *n* = 6, 5 and 24 months, males.

(E) Gene expression in liver was measured by qRT-PCR. Intermediates in the cholesterol biosynthetic pathway are shown on the left and the corresponding genes on the right. Data are normalized to *Myc*^{+/+} for each comparison. *n* = 4, 24 months, males.

(F) Total and nonesterified cholesterol in liver extracts and serum. Normalized to *Myc*^{+/+}. *n* = 5–6, 24 months, males.

(G) Animals of average weight were chosen for rotarod tests, and their performance was corrected for their weight. *n* = 3–4, 24 months, males.

Error bars represent SEM.

See also Figure S4.

understood, varied between tissues: it was unaffected in *Myc*^{+/-} relative to *Myc*^{+/+} mice in heart, reduced in liver and spleen, and increased in lung. *F*₂-isoprostanes, products of lipid peroxidation, are a sensitive and accurate biomarker of oxidative status.

As expected, we found that levels of *F*₂-isoprostanes rose with age, but the changes were the same in *Myc*^{+/+} and *Myc*^{+/-} animals (Figure 5). Hence, *Myc*^{+/-} animals do not seem to be protected from age-associated increases in ROS, or from the

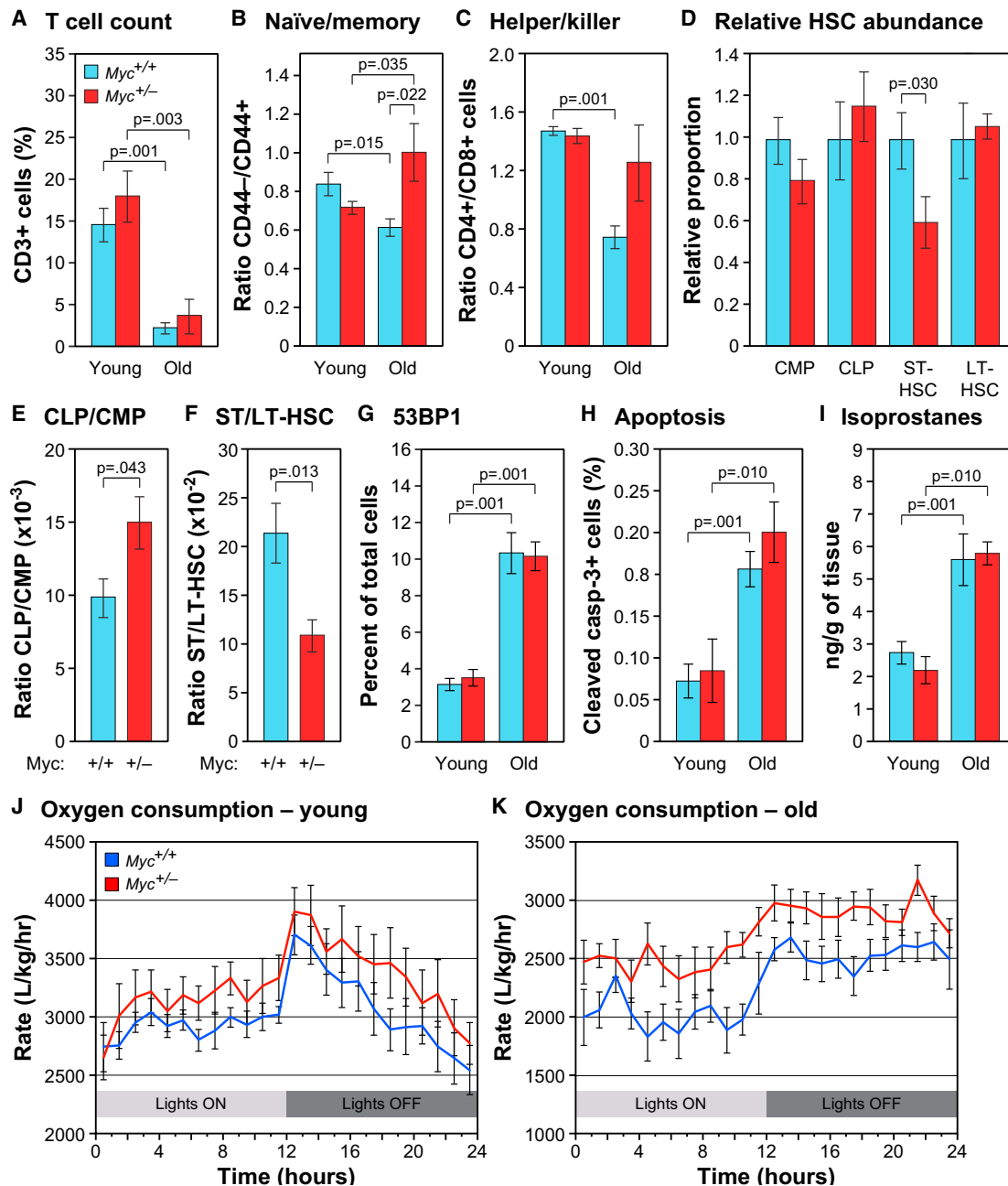


Figure 5. Immunosenescence, Stress Defenses, and Metabolic Activity

(A) Total T cells (CD3⁺) were analyzed by flow cytometry as % of total peripheral lymphocytes. $n = 8-12$, 5 and 24 months, males.

(B and C) Ratio of naïve to memory T cells (CD44⁻/CD44⁺) (B) and ratio of helper to killer T cells (CD4⁺/CD8⁺) (C) were measured in the same samples.

(D) Proportions of common myeloid and lymphoid progenitors (CMP, CLP), and short-term and long-term hematopoietic stem cells (ST-HSC, LT-HSC) were scored as % of Lin⁻ cells in bone marrow (tibia and femur). $n = 6$, 16 months, females. Normalized to $Myc^{+/+}$ for each comparison.

(E and F) Ratio of CLP to CMP (E) and ratio of ST-HSC to LT-HSC (F) in the same samples.

(G) 53BP1-positive cells were visualized by IF in liver sections. $n = 5-6$, 5 and 25 months, males.

(H) Apoptotic cells in liver were identified by IF with an antibody to cleaved caspase-3. $n = 6-8$, 5 and 22 months, females.

(I) F₂ isoprostane levels were measured in liver extracts using gas chromatography and mass spectrometry. $n = 4-7$, 5 and 23-27 months, males.

(J) O₂ consumption by young animals over a 24 hr period. Statistical significance in was computed using two-way ANOVA (time, genotype). Genotype factor was significantly different; $p < 0.001$. $n = 8$, 5 months, both sexes.

(K) O₂ consumption by old animals. $P(\text{genotype}) < 0.001$. $n = 7-8$, 17-22 months, males.

Error bars represent SEM.

See also Figure S5.

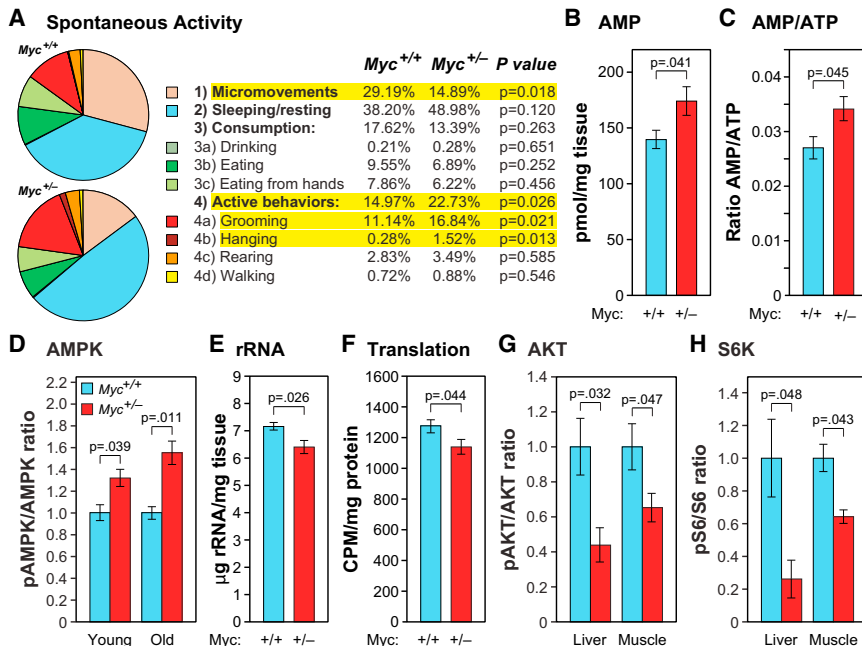


Figure 6. Spontaneous Activity, Energy Metabolism, and Signaling Pathways

(A) Spontaneous home cage activity. Of the four categories of behaviors (micromovements, sleeping, consumption, and active behaviors), micromovements and active behaviors were statistically different. $n = 6$, 16–18 months, males.

(B) AMP concentrations in extracts of muscle. $n = 5$ –7, 25–30 months, both sexes.

(C) AMP to ATP ratio in the same samples (B).

(D) The ratio of phosphorylated (Thr172) to total AMPK α in muscle was determined by immunoblotting. Data are normalized to *Myc*^{+/+} for each comparison. $n = 4$, 9–11 months, females; $n = 3$, 25–30 months, both sexes.

(E) Ribosomal RNA content of liver. $n = 5$, 23–25 months, male.

(F) Translation rates in live animals (liver) were assessed using ³H-phenylalanine incorporation into total protein. Muscle showed the same trend. $n = 5$, 5–7 months; males.

(G) The ratio of phosphorylated (Ser473) to total AKT in liver and muscle. $n = 4$, 9–11 months, female (same samples as in D). Normalized to *Myc*^{+/+}.

(H) The ratio of phosphorylated (Ser235/236) to total S6 ribosomal protein in liver and muscle (same samples as in G). Normalized to *Myc*^{+/+}.

Error bars represent SEM.

See also Figure S6.

consequences of this and other forms of stress, such as DNA damage, apoptosis, and senescence.

Effects on Metabolic Pathways that Regulate Aging

Metabolic rate, measured by O₂ consumption and CO₂ production, declines during normal aging and is increased by CR and in the long-lived Ames and GHR-KO dwarf mice (Bartke and Westbrook, 2012). Interestingly, *Myc*^{+/-} mice have a significantly higher metabolic rate than *Myc*^{+/+} animals (Figures 5J, 5K, S5A, and S5B). Although smaller in magnitude, this trend is already apparent in young *Myc*^{+/-} mice. *Myc*^{+/-} animals also have higher food and water consumption (Figures S5C and S5D). Animals of both genotypes had equivalent body temperatures (Figure S5E). The higher metabolism of *Myc*^{+/-} mice might be needed to thermoregulate their smaller body mass, which would be expected to predispose to hypothermia. A small but significant increase in mitochondrial copy number was found in skeletal muscle (Figure S5F). We also assessed spontaneous activity using a home-cage monitoring system based on automated computer analysis of continuous video recordings (Jhuang et al., 2010) and found that 16- to 18-month-old *Myc*^{+/-} mice displayed a higher level of active behaviors and a lower level of micromovements (Figures 6A and S6A–S6D). Of the active behaviors, hanging from the wire cage roof was notable as being almost completely absent in *Myc*^{+/+} animals. The large difference in this high-energy behavior is consistent with the notably improved rotarod performance (Figure 4G) and lack of osteoporosis (Figures 4B and 4C) in older *Myc*^{+/-} animals.

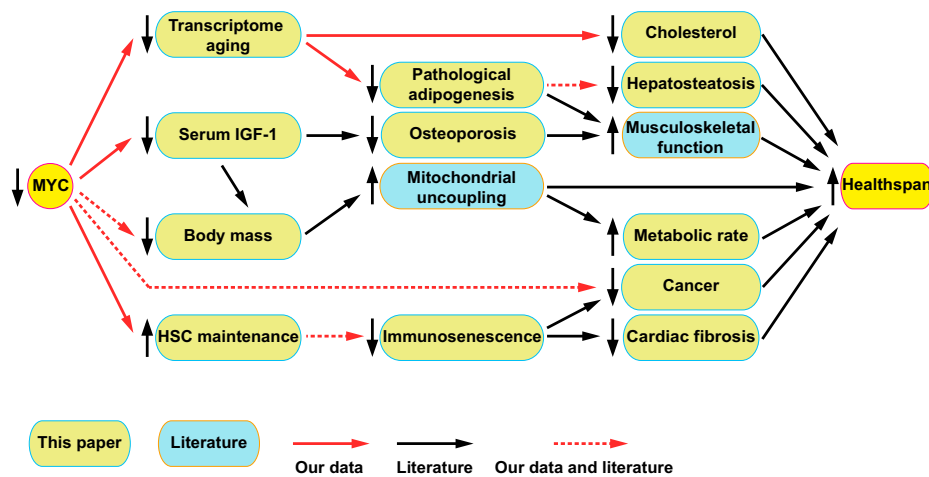
To further investigate these metabolic changes, we measured the levels of the major energy metabolites ATP, ADP, and AMP. We found a small but significant increase in AMP levels and the

AMP to ATP ratio in muscle of old *Myc*^{+/-} mice (Figures 6B and 6C), and the same trends were observed in young animals (Figures S6E and S6F). There was also a trend toward higher ADP levels (Figure S6G) and similar overall trends in liver (Figures S6I and S6J). Although small in magnitude, in aggregate these changes are suggestive of a lower energy status in *Myc*^{+/-} animals, a condition known to activate AMP-dependent kinase (AMPK). In agreement, we observed a significantly higher ratio of Thr172-phosphorylated AMPK α to total AMPK α (Figure 6D), indicative of its activation.

MYC is a major regulator of ribosome biogenesis (Brown et al., 2008). Indeed, analysis of our microarray data suggested that genes encoding ribosomal proteins are downregulated in *Myc*^{+/-} tissues (Figure S6K). We verified this effect by biochemical measurements of rRNA content which showed a small but significant decrease in *Myc*^{+/-} mice (Figure 6E). Finally, to directly query the rate of translation, we measured the in vivo incorporation of a radioactive amino acid into total protein and also found a significant decrease in *Myc*^{+/-} mice (Figure 6F).

Myc^{+/-} mice have reduced levels of circulating IGF-1 (Figure 2C). To more directly examine signaling through this pathway, we assessed the activation status of protein kinase B (AKT). We observed a significantly lower ratio of Ser473-phosphorylated AKT to total AKT in liver and muscle of adult *Myc*^{+/-} mice (Figure 6G), indicative of reduced activity. Both AMPK and AKT regulate mTOR and increased AMPK and decreased AKT activity would promote reduced mTOR activity. We assessed S6 protein kinase activity, a downstream target of mTOR in the protein translation pathway, by measuring the phosphorylation state of ribosomal S6 protein. We found a lower level of phosphorylation in both liver and muscle of adult *Myc*^{+/-} mice (Figure 6H).

A Phenotypes



B Pathways

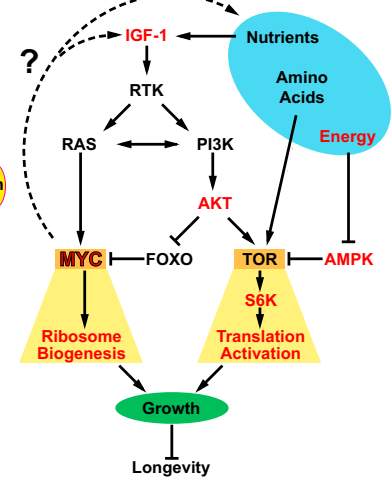


Figure 7. The Effects of *Myc* on Healthspan and Longevity

(A) Phenotypes of *Myc*^{+/-} mice demonstrating their interconnectedness and impact on healthspan (the key under the drawing applies to this panel only).

(B) Pathways affected by *Myc* hypomorphism and their relationship to increased longevity. The components investigated in this report are highlighted in red.

DISCUSSION

Myc^{+/-} mice have significantly extended median and maximum lifespans in both sexes and a reduced mortality rate across all ages relative to their normal *Myc*^{+/+} littermates. *Myc* haploinsufficiency in *Drosophila* also extends lifespan, pointing to a deep conservation of the underlying processes (Greer et al., 2013). *Myc*^{+/-} mice display ameliorated aging phenotypes across a variety of pathophysiological processes in multiple organs, the breadth of which suggests that their increased longevity is not attributable to the prevention of a specific fatal disease, but rather to a broadly increased healthspan.

Aging *Myc*^{+/-} mice have healthier lipid and cholesterol metabolism, stronger bones, less fibrosis, reduced cancer progression, a higher metabolic rate, improved motor control, and reduced immunosenescence compared to their normal littermates. All of these phenotypes oppose the effects of normal aging and are shared with many other long-lived mouse models (Figure 7; Table S7). For example, CR and Ames dwarf mice also have an increased metabolic rate (Bartke and Westbrook, 2012); CR and metformin-treated mice have lower cholesterol (Martin-Montalvo et al., 2013; Tsuchiya et al., 2004); Ames dwarf, CR and rapamycin or metformin administration ameliorate cancer (Ikeno et al., 2003; Martin-Montalvo et al., 2013; Neff et al., 2013); and CR, Ames dwarf, and metformin-treated mice have better motor coordination (Brown-Borg et al., 2012; Lanza et al., 2012; Martin-Montalvo et al., 2013). All of these long-lived models show reduced body mass. However, *Myc* hypomorphic mice also exhibit contrasting phenotypes; for example, rapamycin administration increases serum cholesterol and does not improve motor coordination (Neff et al., 2013), Ames dwarf mice have decreased bone density (Heiman et al., 2003), and CR reduces fecundity (Lee and Longo, 2011). Compared to most other longevity extending interventions, many of which compromise

some aspects of healthy physiology, we have to date not found significant physiological deficits in *Myc*^{+/-} mice.

How can MYC coordinate these effects? It took several decades of work to elucidate the complex spectrum of genes regulated by this transcription factor. In addition to directly binding the promoters of its target genes and upregulating them, MYC also exerts strong secondary effects. For example, it can upregulate the expression of other transcription factors, such as TFAM (Gomes et al., 2013). It can modulate gene expression elicited by other factors, such as ZBTB17 (MIZ1) by directly binding to them (Herkert and Eilers, 2010), or by displacing some factors, such as FOXO3A from their promoters (Peck et al., 2013). Finally, MYC regulates a considerable number of microRNA (miRNA) genes, with widespread effects on gene expression (Dang, 2012).

Many MYC targets, direct or indirect, are involved in anabolic and energy producing processes, and their upregulation is believed to be the major mechanism by which MYC promotes cancer (Dang, 2012). The largest category of genes directly upregulated by MYC are involved in protein translation, and regulation of the ribosome biogenesis (RiBi) regulon is believed to be the core and most ancient function of MYC, extending back to primitive holozoans, the first appearance of the MYC-MAX complex in evolution (Brown et al., 2008). We found that the expression of *Rpl* and *Rps* genes was coordinately reduced in *Myc*^{+/-} tissues. In addition, levels of mature rRNA and in vivo translation rates were decreased. Reducing translation, by a variety of means and in multiple species, is well known to extend lifespan. For example (among many others), yeast and nematodes with mutated ribosomal protein genes, *Drosophila* overexpressing TSC1 or TSC2, and mice fed rapamycin or mutated in S6 kinase 1 are all long-lived (Johnson et al., 2013). Given the major and direct role of MYC on RiBi regulon expression, it is likely that at least part of the longevity of *Myc*^{+/-} mice can be attributed to effects on translation.

TOR is a central hub in the networks that sense nutrients and energy status, and MYC plays a key downstream role in these pathways (Figure 7). For example, in *Drosophila* MYC was reported to mediate the protein biosynthetic capacity in response to insulin signaling (Teleman et al., 2008). Both TOR and MYC are activated by receptor tyrosine kinase signaling, through PI3K, RAS, and AKT, with TOR providing an acute response by posttranslationally regulating its targets and MYC enabling a long-term response through transcriptional regulation. Given the downstream position of MYC, it is remarkable that *Myc*^{+/-} mice display changes throughout this nutrient and energy sensing network, all of which have been documented to promote longevity: increased AMPK activity, decreased AKT and S6K activities, and even reduced circulating levels of IGF-1. None of the genes encoding components of these pathways (including IGF-1 binding proteins) are regulated by MYC in our expression data sets. AMPK is likely to be regulated secondarily by the reduced energy status elicited by the higher metabolic rate and activity of *Myc*^{+/-} animals. Increased AMPK activity would then result in reduced TOR and S6K activities. The PI3K and TOR pathways also contain feedback loops that can be affected by MYC, for example, acute deletion of *Myc* was found to impair TOR signaling during T cell activation (Wang et al., 2011b).

Considering the major longevity models, no set of phenotypes or pathways emerges as a unifying correlation. In other words, extended lifespan can be achieved by amelioration of some age-associated pathophysiologicals even if others are left intact. Aging is increasingly being viewed as a segmental process (Wu et al., 2013), and the phenotypes of *Myc* hypomorphic animals are consistent with this notion. In agreement, a meta-analysis of gene expression showed limited overlap between *Myc*^{+/-} mice and CR, metformin, or resveratrol. The same analysis readily revealed previously reported connections between CR and metformin treatment (Martin-Montalvo et al., 2013) and underscores the apparently unique nature of the *Myc*-elicited lifespan extension.

How can this observation be reconciled with the prominent position of MYC in nutrient sensing pathways? First, while MYC and TOR are both major regulatory hubs, their downstream outputs are only partially overlapping. For example, while TOR activity opposes autophagy (Johnson et al., 2013), MYC upregulates it (Dey et al., 2013). Second, we have not detected significant changes in stress levels, stress management pathways, or stress responses in *Myc*^{+/-} animals. These differences resulted in the low overlap scores returned by the meta-analysis, although these methods clearly detected commonalities in pathways related to metabolism and the immune system.

Production of ROS is well known to increase with age. Although manipulation of these pathways has not yet achieved robust lifespan extension in mammals (Salmon et al., 2010), a stress resistance signature is a component of several important longevity models (CR, reduced somatotrophic signaling, metformin treatment; Table S7). The XME pathway, which promotes the removal of ingested or endogenous toxic metabolites, has been strongly linked with longevity in invertebrate models (Gems and Partridge, 2013) and to a more limited extent in the mouse (Li et al., 2013). Although the expression of some XME genes in liver increased less with age in male *Myc*^{+/-} than *Myc*^{+/+} mice, this

effect was of much smaller magnitude than in other long-lived models. Furthermore, fibroblasts from *Myc*^{+/-} mice did not display increased resistance to a variety of chemical stressors. We looked for changes in the production of ROS (F₂ isoprostanes), as well as some notable consequences of stress exposure (DNA double strand breaks, apoptosis, cellular senescence, p21 and p16 induction). Although we observed the expected age-associated increases, we did not observe any significant or consistent effects of the *Myc*^{+/-} genotype. In aggregate, these results indicate that *Myc*^{+/-} mice can achieve impressive gains in longevity without significant improvements in the efficacy of stress management pathways.

Although it is possible that the phenotypes of *Myc*^{+/-} mice predominantly originate from one cell type or tissue and are mediated in an endocrine manner, given the diverse and widespread nature of these effects we think this is unlikely. Furthermore, cell-autonomous effects are clearly apparent in cell culture. As floxed *Myc* alleles are available (de Alboran et al., 2001; Trumpp et al., 2001), investigating the effects of tissue-specific reduction of MYC activity would be of considerable interest. Similarly, the relationship between lifespan and healthspan benefits and the level of MYC activity is not known. We tested a heterozygous (*Myc*^{+/-}) condition that resulted in ~50% levels of MYC mRNA and protein. Given that *Myc* hypomorphs down to ~25% expression are viable, it would be interesting to ask whether these animals continue to accrue healthspan benefits.

A general theme emerging from aging studies in multiple species is that “less is better”: less food, less somatotrophic/IGF-1 signaling, less anabolic activity, less protein translation, etc., and MYC certainly fits well in this paradigm. The majority of genes studied for their promotion of longevity act in signaling pathways. MYC, as a transcription factor, stands out because of the vast size of its regulome. The *Myc* gene is strongly upregulated by mitogens, and by virtue of its widespread targets constitutes the “gas pedal” that allows a cell to double its mass every 18–24 hr. It is thus not surprising that MYC is necessary during embryonic development, and its deregulated expression strongly promotes cancer. Our discovery that long-term reduction of MYC activity robustly extends lifespan greatly extends our understanding of the biology regulated by this fascinating gene. Given the variety of other physiological benefits, further studies of MYC and its targets should be an interesting avenue to explore in human medicine.

EXPERIMENTAL PROCEDURES

Use and Treatment of Animals

Mice were produced and housed in a specific pathogen-free Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC)-certified barrier facility. All procedures were approved by the Brown University Institutional Animal Care and Use Committees (IACUC) committee. For longevity studies animals, were allowed to die naturally. Both females and males in the longevity study were virgins. Animals of both genotypes and the same sex were housed together.

Statistical and Demographic Analysis

Data are shown as means with SEM (unless stated otherwise). N indicates the number of animals per test group; age and sex are also noted. Student's t test (unpaired, two-tailed, equal variance) was used for all pairwise comparisons. All relevant p values are shown in the figures; if not shown, the values

were >0.05. Demographic data were processed with JMP software to compute mean and median lifespans, SEM, percent increase of the median, and p values (log-rank test) for each cohort. Mortality rate was calculated as log (-log [survival]). Maximum lifespans were calculated as the proportion of each cohort still alive when the total population reached 90% mortality, using Fisher's exact test to determine statistical significance.

Procedures Performed on Live Animals

To score vaginal patency, mice were examined daily from weaning until vaginal opening was observed. Reproductive longevity and output of males were determined by continuously housing young males of either genotype with *Myc*^{+/+} females until pregnancies ceased. Females were replaced every 6 months. For female fecundity, young females of either genotype were continuously mated with one *Myc*^{+/+} male. For rotarod tests the apparatus (MedAssociates) was accelerated continuously from 4 to 40 revolutions per min. O₂ consumption, CO₂ production, food, and water intake were measured using the comprehensive lab animal monitoring system (CLAMS, OxyMax Open Circuit Calorimeter, Columbus Instruments). Spontaneous home cage activity was monitored using fully automated computer vision analysis of continuous video recordings. Wound-healing experiments were performed by introducing one 6 mm diameter full thickness skin punch on the back between the shoulder blades.

Histological Analysis

Mice were euthanized in the morning (8–10 a.m.) by isoflurane anesthesia followed by cervical dislocation. The dissection was performed rapidly (<3 min) by several trained staff members working in concert on one mouse. Paraffin-embedded specimens were stained with H&E. OCT-embedded specimens were used for the determination of fibrosis (Masson's trichrome stain), lipid content (Oil Red O stain), and cellular senescence (senescence-associated β -galactosidase stain). Immunofluorescence microscopy was used to quantify MYC protein, cleaved caspase-3, 53BP1 foci, and macrophages.

Analytical Procedures

Individual gene expression was measured by qRT-PCR (SYBR Green system, ABI 7900 Fast Sequence Detection instrument). Total RNA was extracted from cells or tissue samples (20–50 mg, stored at –80°C) with Trizol reagent (Invitrogen), purified using the RNeasy Mini kit (QIAGEN), and reverse transcribed using the TaqMan kit (Applied Biosystems). Expression profiling was performed using Mouse 1.0 Gene ST arrays (Affymetrix), and analyzed using Expression Console software (Affymetrix) and Ingenuity Pathway Analysis software (Ingenuity Systems). rRNA was purified by affinity capture using the RiboMinus kit (Life Technologies). Chromatin immunoprecipitation was performed with the Magna ChIP kit (Millipore) on nuclei isolated from intact frozen liver tissue. Immunoblotting of proteins was performed using SDS-PAGE of whole cell extracts, electro-transfer onto Immobilon-P membranes (Millipore), and staining with the indicated antibodies. Signals were detected using the LI-COR Odyssey (LI-COR Biosciences) infrared imaging system. Bone density and adipose tissue were assessed using a Scanco Medical Micro-CT 40 instrument on whole animals immediately after euthanasia. Lipids were extracted using the Folch method and cholesterol was measured with the Amplex Red Cholesterol kit (Invitrogen). F₂ isoprostane levels were measured using gas chromatography and mass spectrometry. Leptin and adiponectin were quantified in plasma using the Luminex magnetic bead platform (Millipore) with the mouse adipokine panel and the adiponectin single-plex assay. Free IGF-1 in plasma was measured with an ELISA kit from Abcam. Lymphoid and myeloid cells in blood or bone marrow were quantified by flow cytometry. Blood was harvested by cardiac puncture into heparinized tubes, cells were collected by centrifugation, and erythrocytes were lysed with FACS lysing solution (Becton Dickinson). Bone marrow was flushed out of the femur and tibia. In both cases the cell suspensions were stained with the indicated fluorescently conjugated antibodies and analyzed using a Becton Dickinson FACS Aria instrument. To measure AMP, ADP, and ATP, frozen tissue specimens were homogenized (Fisher Scientific, PowerGen Model 125), extracted in perchloric acid, and neutralized with K₂CO₃ on ice. Ion-paired HPLC was run on an Agilent 1200 series instrument using a ZORBAX Eclipse XDB-C18 column (Agilent). To assess protein translation in live animals, L-³H-phenylal-

anine (Perkin-Elmer, 100–140 Ci/mmol) was combined with unlabeled phenylalanine (135 mM), and injected at 1 ml/100 g body weight into the tail vein. Animals were kept under anesthesia for the entire duration (0–30 min), sacrificed by cervical dislocation, and tissues were snap frozen in liquid nitrogen. Incorporation of ³H-phenylalanine was measured by TCA precipitation of total protein and scintillation counting. Detailed protocols are provided in the [Supplemental Information](#).

ACCESSION NUMBERS

The Gene Expression Omnibus (GEO) accession number for the microarray data reported in this paper is GSE55272.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and eight tables and can be found with this article online <http://dx.doi.org/10.1016/j.cell.2014.12.016>.

AUTHOR CONTRIBUTIONS

J.M.S. conceived the study. J.M.S., J.W.H., X.Z., and M.D.C. designed the experiments. J.W.H., X.Z., M.D.C., A.L.P., L.P., J.M., G.B.H., Y.I., Y.Z., B.F., X.L., T.S., W.Q., K.G.B., and J.M.S. performed experiments and analysis. Y.I., T.S., H.V.R., R.A.M., K.G.B., R.D.C., H.X., N.N., and J.M.S. contributed to supervision of personnel, data interpretation, and writing of the manuscript. J.M.S., J.W.H., and X.Z. wrote the manuscript.

ACKNOWLEDGMENTS

This work was supported by NIH grant R37 AG016694 to J.M.S. J.W.H. was supported in part by grant F30 AG035592. J.M.S. was a Senior Scholar of the Ellison Medical Foundation and a recipient of the Glenn Award for Research on the Biological Mechanisms of Aging. The Brown Molecular Pathology and Genomics core facilities were supported by NIH grants P42 ES013660 and P30 GM103410. R.d.C. is funded by the Intramural Research Program of the National Institute on Aging (NIH).

Received: February 23, 2014

Revised: October 21, 2014

Accepted: December 3, 2014

Published: January 22, 2015

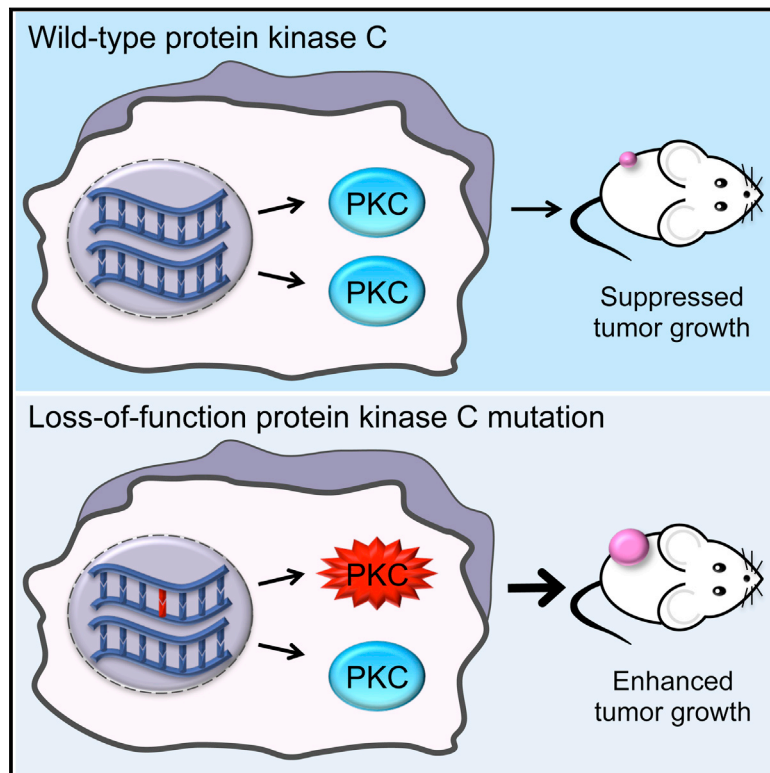
REFERENCES

- Akbar, A.N., and Henson, S.M. (2011). Are senescence and exhaustion intertwined or unrelated processes that compromise immunity? *Nat. Rev. Immunol.* 11, 289–295.
- Anisimov, V.N., Arbeev, K.G., Popovich, I.G., Zabezhinski, M.A., Rosenfeld, S.V., Piskunova, T.S., Arbeeva, L.S., Semchenko, A.V., and Yashin, A.I. (2004). Body weight is not always a good predictor of longevity in mice. *Exp. Gerontol.* 39, 305–319.
- Bartke, A., and Westbrook, R. (2012). Metabolic characteristics of long-lived mice. *Front. Genet.* 3, 288.
- Brown, S.J., Cole, M.D., and Erives, A.J. (2008). Evolution of the holozoan ribosome biogenesis regulon. *BMC Genomics* 9, 442.
- Brown-Borg, H.M., Johnson, W.T., and Rakoczy, S.G. (2012). Expression of oxidative phosphorylation components in mitochondria of long-living Ames dwarf mice. *Age (Dordr.)* 34, 43–57.
- Cho, R.H., Sieburg, H.B., and Muller-Sieburg, C.E. (2008). A new mechanism for the aging of hematopoietic stem cells: aging changes the clonal composition of the stem cell compartment but not individual stem cells. *Blood* 111, 5553–5561.
- Dai, D.F., Santana, L.F., Vermulst, M., Tomazela, D.M., Emond, M.J., MacCoss, M.J., Gollahon, K., Martin, G.M., Loeb, L.A., Ladiges, W.C., and

- Rabinovitch, P.S. (2009). Overexpression of catalase targeted to mitochondria attenuates murine cardiac aging. *Circulation* 119, 2789–2797.
- Dang, C.V. (2012). MYC on the path to cancer. *Cell* 149, 22–35.
- de Alboran, I.M., O'Hagan, R.C., Gärtner, F., Malynn, B., Davidson, L., Rickert, R., Rajewsky, K., DePinho, R.A., and Alt, F.W. (2001). Analysis of C-MYC function in normal cells via conditional gene-targeted mutation. *Immunity* 14, 45–55.
- Dey, S., Tameire, F., and Koumenis, C. (2013). PERK-ing up autophagy during MYC-induced tumorigenesis. *Autophagy* 9, 612–614.
- Eilers, M., and Eisenman, R.N. (2008). Myc's broad reach. *Genes Dev.* 22, 2755–2766.
- Fujimoto, T., and Parton, R.G. (2011). Not just fat: the structure and function of the lipid droplet. *Cold Spring Harb. Perspect. Biol.* 3, 3.
- Gems, D., and Partridge, L. (2013). Genetics of longevity in model organisms: debates and paradigm shifts. *Annu. Rev. Physiol.* 75, 621–644.
- Gomes, A.P., Price, N.L., Ling, A.J., Moslehi, J.J., Montgomery, M.K., Rajman, L., White, J.P., Teodoro, J.S., Wrann, C.D., Hubbard, B.P., et al. (2013). Declining NAD(+) induces a pseudohypoxic state disrupting nuclear-mitochondrial communication during aging. *Cell* 155, 1624–1638.
- Goronzky, J.J., and Weyand, C.M. (2005). T cell development and receptor diversity during aging. *Curr. Opin. Immunol.* 17, 468–475.
- Greer, C., Lee, M., Westerhof, M., Milholland, B., Spokony, R., Vijg, J., and Secombe, J. (2013). Myc-dependent genome instability and lifespan in *Drosophila*. *PLoS ONE* 8, e74641.
- Heiman, M.L., Tinsley, F.C., Mattison, J.A., Hauck, S., and Bartke, A. (2003). Body composition of prolactin-, growth hormone, and thyrotropin-deficient Ames dwarf mice. *Endocrine* 20, 149–154.
- Herkert, B., and Eilers, M. (2010). Transcriptional repression: the dark side of myc. *Genes Cancer* 1, 580–586.
- Hoeijmakers, J.H. (2009). DNA damage, aging, and cancer. *N. Engl. J. Med.* 361, 1475–1485.
- Ikeno, Y., Bronson, R.T., Hubbard, G.B., Lee, S., and Bartke, A. (2003). Delayed occurrence of fatal neoplastic diseases in Ames dwarf mice: correlation to extended longevity. *J. Gerontol. A Biol. Sci. Med. Sci.* 58, 291–296.
- Jang, Y.C., Sinha, M., Cerletti, M., Dall'Osso, C., and Wagers, A.J. (2011). Skeletal muscle stem cells: effects of aging and metabolism on muscle regenerative function. *Cold Spring Harb. Symp. Quant. Biol.* 76, 101–111.
- Jhuang, H., Garrote, E., Mutch, J., Yu, X., Khilnani, V., Poggio, T., Steele, A.D., and Serre, T. (2010). Automated home-cage behavioural phenotyping of mice. *Nat. Commun.* 1, 68.
- Johnson, S.C., Rabinovitch, P.S., and Kaeberlein, M. (2013). mTOR is a key modulator of ageing and age-related disease. *Nature* 493, 338–345.
- Kujoth, G.C., Hiona, A., Pugh, T.D., Someya, S., Panzer, K., Wohlgemuth, S.E., Hofer, T., Seo, A.Y., Sullivan, R., Jobling, W.A., et al. (2005). Mitochondrial DNA mutations, oxidative stress, and apoptosis in mammalian aging. *Science* 309, 481–484.
- Ladiges, W., Van Remmen, H., Strong, R., Ikeno, Y., Treuting, P., Rabinovitch, P., and Richardson, A. (2009). Lifespan extension in genetically modified mice. *Aging Cell* 8, 346–352.
- Lanza, I.R., Zabielski, P., Klaus, K.A., Morse, D.M., Heppelmann, C.J., Bergen, H.R., 3rd, Dasari, S., Walrand, S., Short, K.R., Johnson, M.L., et al. (2012). Chronic caloric restriction preserves mitochondrial function in senescence without increasing mitochondrial biogenesis. *Cell Metab.* 16, 777–788.
- Lee, C., and Longo, V.D. (2011). Fasting vs dietary restriction in cellular protection and cancer treatment: from model organisms to patients. *Oncogene* 30, 3305–3316.
- Li, X., Bartke, A., Berryman, D.E., Funk, K., Kopchick, J.J., List, E.O., Sun, L., and Miller, R.A. (2013). Direct and indirect effects of growth hormone receptor ablation on liver expression of xenobiotic metabolizing genes. *Am. J. Physiol. Endocrinol. Metab.* 305, E942–E950.
- Martin-Montalvo, A., Mercken, E.M., Mitchell, S.J., Palacios, H.H., Mote, P.L., Scheibye-Knudsen, M., Gomes, A.P., Ward, T.M., Minor, R.K., Blouin, M.J., et al. (2013). Metformin improves healthspan and lifespan in mice. *Nat. Commun.* 4, 2192.
- Mateyak, M.K., Obaya, A.J., Adachi, S., and Sedivy, J.M. (1997). Phenotypes of c-Myc-deficient rat fibroblasts isolated by targeted homologous recombination. *Cell Growth Differ.* 8, 1039–1048.
- Matheu, A., Pantoja, C., Efeyan, A., Criado, L.M., Martín-Caballero, J., Flores, J.M., Klatt, P., and Serrano, M. (2004). Increased gene dosage of *Ink4a/Arf* results in cancer resistance and normal aging. *Genes Dev.* 18, 2736–2746.
- Meyer, N., and Penn, L.Z. (2008). Reflecting on 25 years with MYC. *Nat. Rev. Cancer* 8, 976–990.
- Neff, F., Flores-Dominguez, D., Ryan, D.P., Horsch, M., Schröder, S., Adler, T., Afonso, L.C., Aguilar-Pimentel, J.A., Becker, L., Garrett, L., et al. (2013). Rapamycin extends murine lifespan but has limited effects on aging. *J. Clin. Invest.* 123, 3272–3291.
- Pearson, K.J., Baur, J.A., Lewis, K.N., Peshkin, L., Price, N.L., Labinskyy, N., Swindell, W.R., Kamara, D., Minor, R.K., Perez, E., et al. (2008). Resveratrol delays age-related deterioration and mimics transcriptional aspects of dietary restriction without extending life span. *Cell Metab.* 8, 157–168.
- Peck, B., Ferber, E.C., and Schulze, A. (2013). Antagonism between FOXO and MYC regulates cellular powerhouse. *Front. Oncol.* 3, 96.
- Salmon, A.B., Richardson, A., and Pérez, V.I. (2010). Update on the oxidative stress theory of aging: does oxidative stress play a role in aging or healthy aging? *Free Radic. Biol. Med.* 48, 642–655.
- Syed, F.A., and Melim, T. (2011). Rodent models of aging bone: an update. *Curr. Osteoporos. Rep.* 9, 219–228.
- Teleman, A.A., Hietakangas, V., Sayadian, A.C., and Cohen, S.M. (2008). Nutritional control of protein biosynthetic capacity by insulin via Myc in *Drosophila*. *Cell Metab.* 7, 21–32.
- Trumpp, A., Refaeli, Y., Oskarsson, T., Gasser, S., Murphy, M., Martin, G.R., and Bishop, J.M. (2001). c-Myc regulates mammalian body size by controlling cell number but not cell size. *Nature* 414, 768–773.
- Tsuchiya, T., Dhahbi, J.M., Cui, X., Mote, P.L., Bartke, A., and Spindler, S.R. (2004). Additive regulation of hepatic gene expression by dwarfism and caloric restriction. *Physiol. Genomics* 17, 307–315.
- Vafa, O., Wade, M., Kern, S., Beeche, M., Pandita, T.K., Hampton, G.M., and Wahl, G.M. (2002). c-Myc can induce DNA damage, increase reactive oxygen species, and mitigate p53 function: a mechanism for oncogene-induced genetic instability. *Mol. Cell* 9, 1031–1044.
- Wang, J., Geiger, H., and Rudolph, K.L. (2011a). Immunoaging induced by hematopoietic stem cell aging. *Curr. Opin. Immunol.* 23, 532–536.
- Wang, R., Dillon, C.P., Shi, L.Z., Milasta, S., Carter, R., Finkelstein, D., McCormick, L.L., Fitzgerald, P., Chi, H., Munger, J., and Green, D.R. (2011b). The transcription factor Myc controls metabolic reprogramming upon T lymphocyte activation. *Immunity* 35, 871–882.
- Whitfield, J.R., and Soucek, L. (2012). Tumor microenvironment: becoming sick of Myc. *Cell. Mol. Life Sci.* 69, 931–934.
- Wu, J.J., Liu, J., Chen, E.B., Wang, J.J., Cao, L., Narayan, N., Fergusson, M.M., Rovira, I.I., Allen, M., Springer, D.A., et al. (2013). Increased mammalian lifespan and a segmental and tissue-specific slowing of aging after genetic reduction of mTOR expression. *Cell Rep.* 4, 913–920.
- Yuan, R., Tsaih, S.W., Petkova, S.B., Marin de Evsikova, C., Xing, S., Marion, M.A., Bogue, M.A., Mills, K.D., Peters, L.L., Bult, C.J., et al. (2009). Aging in inbred strains of mice: study design and interim report on median lifespans and circulating IGF1 levels. *Aging Cell* 8, 277–287.
- Zahn, J.M., Poosala, S., Owen, A.B., Ingram, D.K., Lustig, A., Carter, A., Weeraratna, A.T., Taub, D.D., Gorospe, M., Mazan-Mamczarz, K., et al. (2007). AGEMAP: a gene expression database for aging in mice. *PLoS Genet.* 3, e201.

Cancer-Associated Protein Kinase C Mutations Reveal Kinase's Role as Tumor Suppressor

Graphical Abstract



Authors

Corina E. Antal, Andrew M. Hudson, ..., John Brognard, Alexandra C. Newton

Correspondence

john.brognard@cruk.manchester.ac.uk (J.B.),
anewton@ucsd.edu (A.C.N.)

In Brief

Cancer-associated kinase mutations have generally been characterized as oncogenic, but an analysis of PKC mutations reveals that the majority are loss of function, indicating a tumor-suppressive role for this kinase and a shift in therapeutic strategies targeting PKC.

Highlights

- Cancer-associated PKC mutations are LOF and can act in a dominant-negative manner
- Correcting a heterozygous PKC β LOF mutation reduces tumor volume
- Hemizygous deletion shows PKC is haploinsufficient for tumor suppression
- Therapeutic strategies should aim to restore PKC activity instead of inhibiting it



Cancer-Associated Protein Kinase C Mutations Reveal Kinase's Role as Tumor Suppressor

Corina E. Antal,^{1,2} Andrew M. Hudson,³ Emily Kang,¹ Ciro Zanca,⁴ Christopher Wirth,⁵ Natalie L. Stephenson,³ Eleanor W. Trotter,³ Lisa L. Gallegos,^{1,2,7} Crispin J. Miller,⁵ Frank B. Furnari,⁴ Tony Hunter,⁶ John Brognard,^{3,*} and Alexandra C. Newton^{1,*}

¹Department of Pharmacology, University of California at San Diego, La Jolla, CA 92093, USA

²Biomedical Sciences Graduate Program, University of California at San Diego, La Jolla, CA 92093, USA

³Signalling Networks in Cancer Group, Cancer Research UK Manchester Institute, University of Manchester, Manchester M20 4BX, UK

⁴Ludwig Institute for Cancer Research, University of California at San Diego, La Jolla, CA 92093, USA

⁵Applied Computational Biology and Bioinformatics Group, Cancer Research UK Manchester Institute, University of Manchester, Manchester M20 4BX, UK

⁶The Salk Institute, La Jolla, CA 92037, USA

⁷Present address: Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA

*Correspondence: john.brognard@cruk.manchester.ac.uk (J.B.), anewton@ucsd.edu (A.C.N.)

<http://dx.doi.org/10.1016/j.cell.2015.01.001>

SUMMARY

Protein kinase C (PKC) isoforms have remained elusive cancer targets despite the unambiguous tumor promoting function of their potent ligands, phorbol esters, and the prevalence of their mutations. We analyzed 8% of PKC mutations identified in human cancers and found that, surprisingly, most were loss of function and none were activating. Loss-of-function mutations occurred in all PKC subgroups and impeded second-messenger binding, phosphorylation, or catalysis. Correction of a loss-of-function PKC β mutation by CRISPR-mediated genome editing in a patient-derived colon cancer cell line suppressed anchorage-independent growth and reduced tumor growth in a xenograft model. Hemizygous deletion promoted anchorage-independent growth, revealing that PKC β is haploinsufficient for tumor suppression. Several mutations were dominant negative, suppressing global PKC signaling output, and bioinformatic analysis suggested that PKC mutations cooperate with co-occurring mutations in cancer drivers. These data establish that PKC isoforms generally function as tumor suppressors, indicating that therapies should focus on restoring, not inhibiting, PKC activity.

INTRODUCTION

The protein kinase C (PKC) family has been intensely investigated in the context of cancer since the discovery that it is a receptor for the tumor-promoting phorbol esters (Castagna et al., 1982). This led to the dogma that activation of PKC by phorbol esters promotes carcinogen-induced tumorigenesis (Griner and Kazanietz, 2007), yet targeting PKC in cancer has been unsuccessful.

The PKC family contains nine genes that have many targets and thus diverse cellular functions, including cell survival, prolif-

eration, apoptosis, and migration (Dempsey et al., 2000). PKC isoforms comprise three classes: conventional (cPKC: α , β , γ), novel (nPKC: δ , ϵ , η , θ), and atypical (aPKC: ζ , ι). cPKC and nPKC isoforms are constitutively phosphorylated at three priming sites (activation loop, turn motif, and hydrophobic motif) to structure PKC for catalysis (Newton, 2003). A pseudosubstrate segment maintains PKC in an autoinhibited conformation that is relieved by second-messenger binding. cPKC isoforms are activated by binding to diacylglycerol (DAG) and Ca^{2+} , whereas nPKC isoforms are activated solely by DAG, events that engage PKC at membranes. Thus, these PKC isoforms have two prerequisites for activation: constitutive processing phosphorylations and second-messenger-dependent relocalization to membranes. Prolonged activation of cPKC and nPKC isoforms with phorbol esters leads to their dephosphorylation and subsequent degradation, a process referred to as downregulation (Hansra et al., 1996; Young et al., 1987). aPKC isoforms bind neither Ca^{2+} nor DAG.

PKC has proved an intractable target in cancer therapeutics (Kang, 2014). PKC ι was proposed to be an oncogene in lung and ovarian cancers (Justilien et al., 2014; Regala et al., 2005; Zhang et al., 2006), and PKC ϵ was categorized as an oncogene because of its ability to transform cells (Cacace et al., 1993). However, for most PKC isoforms, there is conflicting evidence as to whether they act as oncogenes or as tumor suppressors. For example, PKC δ is considered a tumor suppressor because of its pro-apoptotic effects (Reyland, 2007). However, it promotes tumor progression of lung and pancreatic cancers in certain contexts (Mauro et al., 2010; Symonds et al., 2011). Similarly, both overexpression and loss of PKC ζ in colon cancer cells have been reported to decrease tumorigenicity in nude mice or cell lines, respectively (Luna-Ulloa et al., 2011; Ma et al., 2013). Likewise, PKC α was reported to both induce (Walsh et al., 2004; Wu et al., 2013) and suppress colon cancer cell proliferation (Gwak et al., 2009) and to suppress colon tumor formation in the APC^{Min/+} model (Oster and Leitges, 2006). Based on the dogma that PKC isoforms contribute positively to cancer progression, many PKC inhibitors have entered clinical trials; however, they have been ineffective (Mackay and Twelves, 2007).

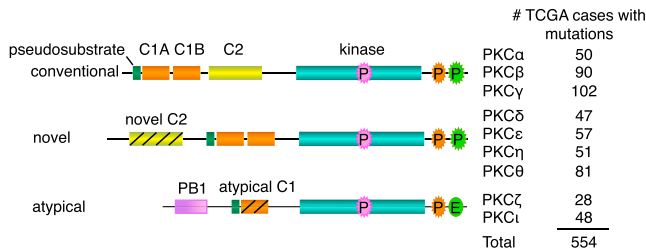


Figure 1. A Multitude of Cancer-Associated Mutations Have Been Identified within the Nine PKC Genes

(Left) Domain structure of conventional (α , β , γ), novel (δ , ϵ , η , θ), and atypical (ζ , ι) PKC members showing priming phosphorylation sites: activation loop (pink), turn motif (orange), and hydrophobic motif (green). (Right) Number of TCGA cases with cancer-associated mutations (missense, nonsense, insertions, deletions, splice site, or translation start site) identified within each of the PKC genes.

In fact, a recent meta-analysis of controlled trials of PKC inhibitors combined with chemotherapy versus chemotherapy alone revealed that PKC inhibitors significantly decreased response rates and disease control rates in non-small cell lung cancer (Zhang et al., 2014). Why has inhibiting PKC failed in the clinic? It has been well established that prolonged or repetitive treatment with phorbol esters depletes cPKC and nPKC isozymes from cells (Blumberg, 1980; Nelson and Alkon, 2009), bringing into question whether loss of PKC, rather than its activation, promotes tumorigenesis.

PKC is frequently mutated in human cancers. To uncover whether loss or gain of PKC function contributes to cancer progression, we selected mutations throughout the primary sequence and family membership and assessed their functional impact. Specifically, we asked how these cancer-associated mutations alter the signaling output of PKC using our genetically encoded reporter, C kinase activity reporter (CKAR) (Violin et al., 2003). Characterization of 46 of these mutations revealed that most reduced or abolished PKC activity and none were activating. Bioinformatic analysis of all PKC mutations revealed that they may cooperate with co-occurring mutations in oncogenes and tumor suppressors known to be regulated by PKC. Correction of one patient-identified, heterozygous, loss-of-function (LOF) PKC β mutation in a colon cancer cell line significantly decreased tumor size in mouse xenografts, indicating that loss of PKC function enhances tumor growth. Our data are consistent with PKC isozymes functioning generally as tumor suppressors, reversing the paradigm that their hyperactivation promotes tumor growth.

RESULTS

A Multitude of Cancer-Associated Mutations Have Been Identified within the Nine PKC Genes

554 mutations (as of October 2014), of which most are heterozygous, have been identified in diverse cancers (Cerami et al., 2012; Gao et al., 2013) within cPKC (242), nPKC (236), and aPKC (76) isozymes (Figure 1). These mutations reside throughout the entire coding region, with no apparent mutational hot-spots. Therefore, we conducted a comprehensive study of

mutations within PKC domains and within interdomain regions to determine how they affect PKC signaling to contribute to cancer pathogenesis. 46 mutations of both conserved and non-conserved residues were selected from all three classes of PKC isozymes (Table 1 and Table S1).

PKC Mutations in the Regulatory C1 and C2 Domains Are LOF

The C1 domains of cPKC and nPKC isozymes are critical for their activation because they mediate PKC translocation to membranes via binding to DAG. Thus, we investigated how C1 domain mutations alter PKC translocation and activation. To measure agonist-dependent PKC activity, COS7 cells co-expressing the FRET-based PKC reporter (CKAR) and equal levels of either wild-type (WT) or mutant mCherry-tagged PKC were stimulated with the cell-permeable DAG, DiC8, or the phorbol ester, phorbol 12,13-dibutyrate (PDBu), and phosphorylation-dependent FRET ratio changes were recorded. Phorbol esters serve as an effective although non-physiological tool to maximally activate PKC because they bind with 100-fold higher affinity to C1 domains compared to DAG (Mosior and Newton, 1998). A mutation identified in a colorectal cancer tumor altered a residue (PKC α H75Q) required for coordination of Zn²⁺ and thus for folding of the C1 domain (Figure 2A). This mutation ablated agonist-stimulated activity, as evidenced by a lower FRET ratio trace compared with that of cells containing only endogenous PKC (Figure 2B). This lower activity suggests that the mutant is dominant negative toward global PKC output. Within a head and neck cancer patient, a mutation altered a critical residue (PKC α W58L) required for controlling the affinity for DAG, but not phorbol ester (Dries et al., 2007) (Figure 2A). This mutation also abolished DiC8-induced and basal activity but retained some PDBu-induced activity, consistent with this residue selectively regulating DAG affinity (Figures 2B and S1A). Because membrane translocation is a prerequisite for activation of cPKC isozymes, we compared the translocation of YFP-tagged WT and mutant PKC to membrane-targeted CFP using FRET (Antal et al., 2014). Mutation of either residue impaired translocation upon stimulation with DiC8, phorbol ester (Figure 2C), or the natural agonist UTP (Figure 2D), accounting for the inability of these agonists to activate the mutants. Lastly, we asked how these mutations affected the processing phosphorylations of PKC. PKC α H75Q, but not W58L, was unphosphorylated, likely because the misfolded C1A domain of the H75Q mutant prevented its processing (Figure 2E). Three additional mutations within the C1A domains of PKC α (G61W), PKC β (G61W), and PKC γ (Q62H) also exhibited reduced agonist-induced PKC activity (Figures S1B–S1D). Our analysis of nine C1 domain mutations revealed that five reduced or abolished activity while none were hyperactivating (Tables 1 and S1). Inactivation occurred by altering two key inputs required for PKC function: disruption of binding to DAG or processing by phosphorylations.

The C2 domain of cPKC isozymes is also critical for activation, as it mediates Ca²⁺-dependent pre-targeting to plasma membrane, where these isozymes bind DAG and become activated (Newton, 2003). One mutation identified within the C2 domain of PKC γ (D193N) was present in colorectal and ovarian cancers

Table 1. Loss-of-Function PKC Mutations in Cancer

Mutation ^a	Activity	Domain	Cancer(s)	Residue Importance	Allele Frequency	Other Mutations ^b
γ G23E	none ^c	PS	colorectal	adding negative charge to pseudosubstrate	N/A	γ G23W δ G146R ι G128C
ε R162H	low		head and neck	non-conserved	0.15	
α W58L	none ^c	C1A	head and neck	DAG binding; conserved in all C1a domains	0.22	γ W57splice θ W171*
α G61W	low		lung	conserved in cPKC C1a domains	0.05	β G61W
β G61W	low		lung	conserved in cPKC C1a domains	0.06	α G61W
γ Q62H	none ^c		lung	conserved in all PKC isoforms	0.45	α Q63H ε Q197P
α H75Q	none ^d		colorectal	coordinates Zn ²⁺ ; conserved in all C1 domains	N/A	η H284Y ι H179Y
γ D193N	none ^c	C2	colorectal/melanoma/ ovarian	Ca ²⁺ binding site	0.28	
γ T218M	none ^c		stomach	non-conserved	0.42	γ T218R
γ D254N	low		endometrial/ovarian	Ca ²⁺ binding site	0.43	
α G257V	none ^c		lung	conserved in cPKC isoforms	0.12	
γ F362L	none ^c	Kinase	endometrial	conserved in cPKC and nPKC isoforms	0.21	γ F362fs β F353L
β Y417H	none ^c		liver	conserved in cPKC isoforms	0.67	γ Y431F
ζ E421K	none ^d		breast	APE motif; conserved in most protein kinases	N/A	α E508K ι E423D
α F435C	none ^c		endometrial	conserved in cPKC and nPKC isoforms	0.31	
α A444V	low		endometrial/breast	conserved in cPKC and nPKC isoforms	0.27	β A447T γ A461T γ A461V δ A454V θ A485T ι S359C
γ G450C	none ^c		endometrial/lung/liver	conserved in cPKC isoforms	0.41	ε R502*
α D481E	low		colorectal	DFG motif; conserved in most protein kinases	N/A	β D484N γ D498N ι D396E
β A509V	none ^d		breast	APE motif; conserved in most protein kinases	N/A	α A506V α A506T β A509T
β A509T	none ^c		colorectal	APE motif; conserved in most protein kinases	0.53	α A506V α A506T β A509V
γ P524R	none ^d		pancreatic	APE motif; conserved in most protein kinases	N/A	γ P524L δ P517S ε P576S θ P548S
δ D530G	none ^d		colorectal	anchors the conserved regulatory spine; conserved in all eukaryotic kinases	N/A	β D523N γ D537G γ D537Y
δ P568A	none ^c		head and neck	conserved in all PKC isoforms	0.16	δ P568S β P561H γ P575H
β G585S	low		lung	conserved in all PKC isoforms	N/A	η G598V
η K591E	low		breast	reversal of conserved charge	N/A	η K591N θ R616Q
η R596H	none ^d		colorectal	conserved in all PKC isoforms	0.50	
η G598V	none ^d		lung	conserved in all PKC isoforms	N/A	β G585S

(Continued on next page)

Table 1. Continued

Mutation ^a	Activity	Domain	Cancer(s)	Residue Importance	Allele Frequency	Other Mutations ^b
β P619Q	none ^d	C-tail	endometrial	PXXP motif; conserved in AGC kinases	0.48	

PKC mutations showing no activity with any agonist, no activity with physiological stimuli, or reduced activity in response to physiological stimuli. Allele frequencies were obtained from cBioPortal.

^aMutations examined in this study.

^bOther mutations present at the same/corresponding residue in the same/other PKC isozymes.

^cKinase-dead.

^dNo response to physiological stimuli.

and in melanoma. Another (D254N) was found in endometrial and ovarian cancers. Because both of these Asp residues (Figure 2F) coordinate Ca^{2+} (Medkova and Cho, 1998), we monitored their activation upon elevation of intracellular Ca^{2+} with thapsigargin, a sarco/endoplasmic reticulum Ca^{2+} -ATPase inhibitor (Rogers et al., 1995). In contrast to WT PKC γ , neither mutant was activated (Figure 2G) nor translocated to the plasma membrane (Figure 2H) following thapsigargin addition, consistent with impaired Ca^{2+} binding. However, both mutants retained full responses to phorbol esters, consistent with unimpaired C1 domains. To further substantiate the inability of the mutants to bind Ca^{2+} , we monitored PKC oscillatory translocation stimulated by histamine-induced oscillatory Ca^{2+} release in HeLa cells (Violin et al., 2003). Whereas WT PKC γ exhibited oscillatory translocation in some cells, the C2 domain mutants were unresponsive to histamine (Figure 2I). Thus, these C2 domain mutations dampen PKC γ activity because they impede Ca^{2+} binding. Mutation of two other C2 domain residues that are not directly involved in Ca^{2+} binding (PKC γ T218M and PKC α G257V) also caused LOF (Figure S1D and S1E); PKC α G257V was LOF because it was not processed by phosphorylation (Figure S1F), whereas the remaining C2 domain mutants were (data not shown). Our analysis of six C2 domain mutations revealed four LOF mutations and no hyperactivating ones (Tables 1 and S1).

PKC Mutations in the Kinase Domain Are LOF

We next evaluated 21 kinase domain mutations, two of which were within PKC δ : D530G in colorectal cancer and P568A in head and neck cancer (Figure 3A). Asp530 functions as an anchor for the kinase regulatory spine, a highly conserved structural element of eukaryotic kinases (Kornev et al., 2006; Kornev et al., 2008); not surprisingly, the D530G mutant was kinase dead and not primed by phosphorylation (Figures 3B and 3C). Mutation of the conserved Pro568 to Ala also prevented a response to natural agonist stimulation but maintained some PDBu-stimulated activity, likely because a small pool of this mutant was phosphorylated (Figures 3B and 3C).

Strikingly, all three PKC η mutations examined (K591E, R596H, and G598V) altered its subcellular localization by pre-localizing it at the plasma membrane prior to stimulation (Figure 3D). However, despite constitutive membrane association, these mutants had reduced basal and stimulated activity as read out by a phospho-(Ser) PKC substrate antibody (Figure 3E) because they were not processed by phosphorylation (Figure 3F). We have previously shown that unprocessed nPKC isozymes have exposed C1 domains that induce constitutive membrane association (Antal et al., 2014).

A number of mutations were present within the highly conserved APE motif that is involved in substrate binding and allosteric activation of kinases (Kornev et al., 2008). PKC γ P524R and PKC β A509V mutations ablated activity by preventing processing phosphorylations, and both exhibited dominant-negative roles (Figures 3G–3J). PKC β A509T (colorectal cancer) also showed loss of function in response to UTP but was modestly activated by the potent ligand PDBu (Figure 3I), likely because a small pool of it was phosphorylated (Figure 3J). A LOF mutation that prevented processing of the atypical PKC ζ was also found within the APE motif (E421K; Figure S1G).

Further analysis revealed that 16 out of 21 kinase domain mutations that we analyzed (Tables 1 and S1) resulted in full or partial LOF, with the majority preventing processing by phosphorylation. For example, PKC α F435C, PKC α A444V, PKC β II Y417H, PKC β II G585S, and PKC γ G450C had impaired phosphorylation and reduced activity (Figures S1C–S1F and S1H–S1J). However, partial LOF mutations were also observed in cases in which phosphorylation was maintained—PKC α D481E (Figures S1B and S1F) and PKC γ F362L (Figures S1D and S1J), suggesting that these mutations likely decrease PKC's intrinsic catalytic activity.

The Majority of Cancer-Associated PKC Mutations Are LOF

Our analysis of 46 mutations present within eight of the PKC genes revealed that ~61% (28) of them were LOF and none were activating (Figure 4A). A lack of identification of activating mutations is not an artifact of our assays, as activating PKC mutations that increase PKC affinity for DAG or decrease autoinhibition are readily detectable (data not shown). LOF mutations were identified within cPKC (α , β , γ), nPKC (δ , ϵ , η), and aPKC (ζ) isozymes and occurred within the C1, C2, and kinase domains as well as the pseudosubstrate and C-terminal tail (Figure 4B). For example, the PKC γ G23E pseudosubstrate mutation was not processed by phosphorylation (Figure S1J) and thus lacked any UTP-stimulated activity (Figure S1D), and the PKC ϵ R162H pseudosubstrate mutation showed reduced agonist-stimulated and basal activity (Figures S1K and S1L). The PKC β P619Q C-terminal tail mutation, residing within a conserved PXXP motif required for processing (Gould et al., 2009), was also LOF as it prevented PKC phosphorylation (Figure S1H). Overall, PKC LOF occurred by diverse mechanisms, most commonly by preventing processing phosphorylations or ligand binding, and as such, there were no mutational hotspots for loss of function. However, we identified seven LOF mutation “warmspots” (Sun et al., 2007) that fell within highly conserved regions of PKC—one within the pseudo-substrate and six within the kinase domain (Figure 4C). Thus,

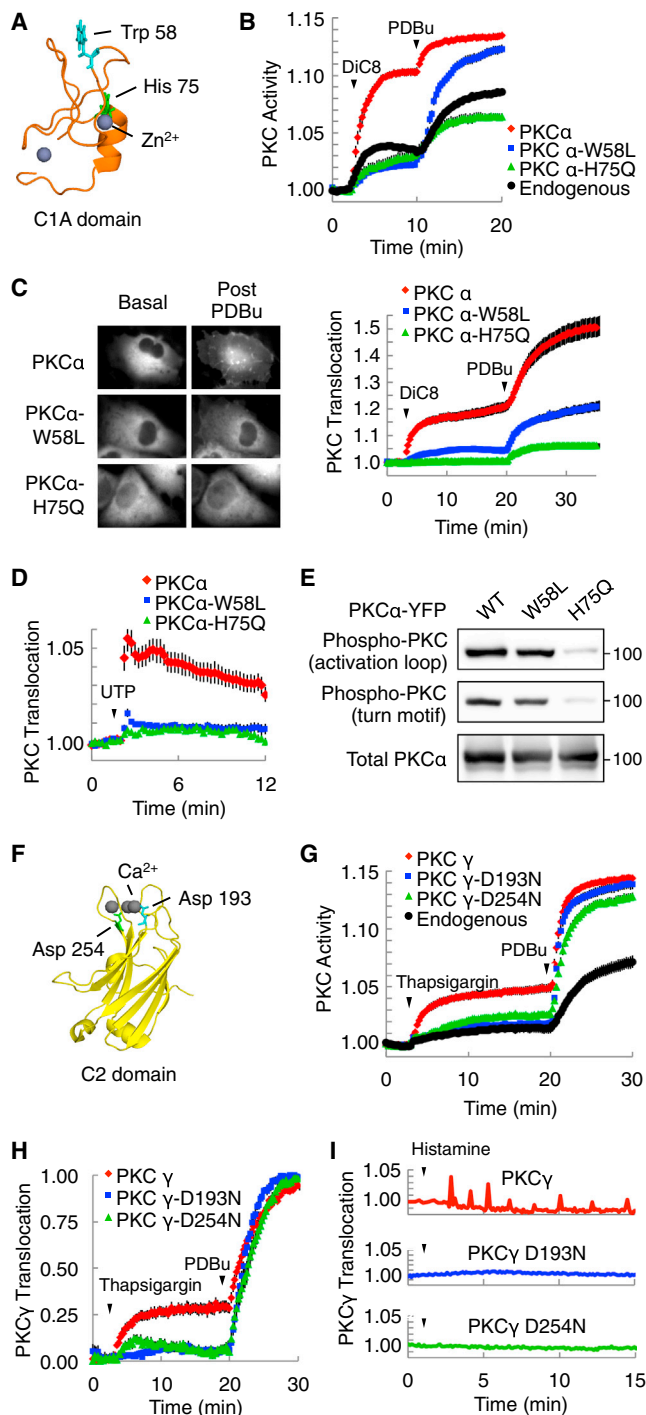


Figure 2. PKC Mutations in the Regulatory C1 and C2 Domains Are LOF

(A) Solution structure of the C1A domain of PKC γ (PDB 2E73) showing the corresponding PKC α His75 residue that coordinates Zn²⁺ and PKC α Trp58. (B) Normalized FRET ratio changes (mean \pm SEM) representing DiC8- (10 μ M) followed by PDBu- (200 nM) induced PKC activity as read out by CKAR in COS7 cells co-expressing CKAR and either mCherry-tagged WT, mutant PKC α , or no exogenous PKC (endogenous). (C) (Left) Representative YFP images of the indicated PKC isoforms under basal and PDBu-treated conditions (200 nM; 15 min) showing relocalization of

inactivating mutations targeted conserved regulatory elements and frequently hit the same residue, whereas mutations that exhibited no difference from WT occurred more randomly (Table S1).

Analysis of cancer types most frequently harboring PKC mutations revealed that, although PKC isoforms are mutated across many cancers, PKC mutations are enriched in certain cancers (Figure 4D). Namely, PKC isoforms are mutated in 20%–25% of melanomas, colorectal cancers, or lung squamous cell carcinomas but are mutated in <5% of ovarian cancers, glioblastoma, or breast cancers (Cerami et al., 2012; Gao et al., 2013). Additionally, nPKC isoforms are most commonly mutated in gastrointestinal cancers (pancreatic, stomach, and colorectal), which have a lower mutation burden than melanomas and lung cancers, highlighting their importance in this type of cancer (Figure 4D). The majority of PKC mutations are heterozygous, with an allele frequency varying from 0.05 to 0.67 for the mutations characterized (Tables 1 and S1). This indicates that PKC mutations can be truncal events in regards to tumor heterogeneity and exist in a majority of the cells within a tumor or can be branchal events acquired later in tumorigenesis as the tumor progresses to a more aggressive stage. This is consistent with PKC mutations being co-driver events that enhance tumorigenesis mediated by primary drivers.

Dominant-Negative PKC β Mutation Confers a Tumor Growth Advantage

Because the majority of PKC mutations examined were LOF, we tested whether we could rescue HCT116 colon cancer cells that have a heterozygous LOF frameshift mutation in the C2 domain of PKC β by overexpressing WT PKC β II. This resulted in a dramatic reduction in anchorage-independent growth (Figure S2A), a hallmark of cellular transformation. Thus, we next used CRISPR/Cas9-mediated genome editing to ask whether

WT, but not mutant PKC α , to membranes. (Right) Normalized FRET ratio changes (mean \pm SEM) quantifying translocation of YFP-tagged PKC α proteins toward a membrane-targeted CFP upon stimulation with 10 μ M DiC8, followed by 200 nM PDBu.

(D) Normalized FRET ratio changes (mean \pm SEM) showing PKC translocation following UTP (100 μ M) stimulation.

(E) Immunoblot showing the phosphorylation state of the indicated YFP-tagged PKC α proteins.

(F) Crystal structure of the C2 domain of PKC γ (PDB 2UZP) highlighting Asp193 and Asp254 residues involved in Ca²⁺ binding.

(G) Normalized FRET ratio changes (mean \pm SEM) showing PKC activity as read out by CKAR upon elevation of intracellular Ca²⁺ stimulated by thapsigargin (5 μ M), followed by PDBu (200 nM).

(H) Normalized FRET ratio changes (mean \pm SEM) showing translocation of YFP-tagged PKC γ constructs toward membrane-localized CFP upon stimulation of COS7 cells with thapsigargin (5 μ M) followed by PDBu (200 nM). Data were normalized to the maximal amplitude of translocation for each cell and then scaled from 0 to 1 using the equation: $X = (Y - Y_{min}) / (Y_{max} - Y_{min})$, where Y = normalized FRET ratio, Y_{min} = minimum value of Y , and Y_{max} is maximum value of Y .

(I) Normalized FRET ratio changes displaying oscillatory translocation of YFP-tagged WT PKC γ , but not PKC γ mutants D193N and D254N, in HeLa cells co-expressing membrane-targeted CFP and stimulated with 10 μ M histamine. Data are representative traces from individual cells of three independent experiments.

See also Figure S1.

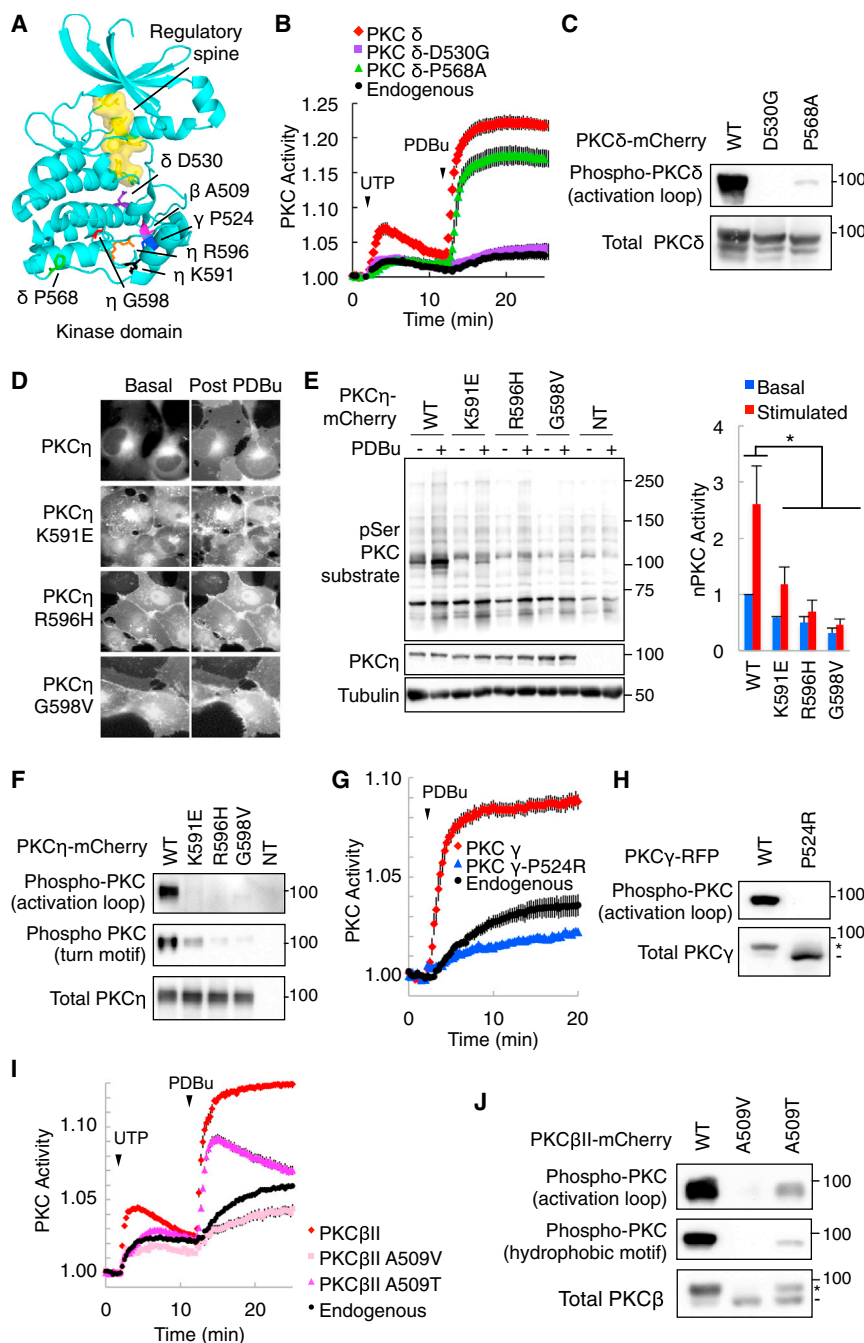


Figure 3. PKC Mutations in the Kinase Domain Are LOF

(A) Crystal structure of the kinase domain of PKC δ III (PDB 210E) highlighting cancer-associated residues and the regulatory spine (yellow space filling).

(B) Normalized FRET ratio changes (mean \pm SEM) showing PKC activity of PKC δ constructs in COS7 cells co-expressing the plasma membrane-targeted, PKC δ -specific reporter PM- δ KAR. Cells were stimulated with UTP (100 μ M) followed by PDBu (200 nM).

(C) Immunoblot analysis of the phosphorylation state of PKC δ WT and mutants.

(D) Representative mCherry images of mCherry-tagged PKC η WT or mutants showing localization under basal conditions and 15 min post 200 nM PDBu addition to COS7 cells.

(E) (Left) Immunoblot showing PKC substrate phosphorylation. COS7 cells overexpressing the indicated constructs were pre-treated with 4 μ M Gö6976 for 10 min to inhibit cPKC isozymes and were then stimulated or not with 200 nM PDBu to activate nPKC isozymes. (Right) Immunoblots were quantified and normalized to total PKC η levels and tubulin. Data represent averages of three independent experiments \pm SEM. Comparisons for basal and stimulated activity were made using a repeated-measures one-way ANOVA followed by post hoc Dunnett's multiple comparison test. * $p < 0.05$ as compared with the WT group.

(F) Immunoblot analysis of the phosphorylation state of mCherry-tagged PKC η WT and mutants.

(G) Normalized FRET ratio changes (mean \pm SEM) showing PKC activity from COS7 cells co-expressing CKAR and RFP-tagged PKC γ mutants stimulated with 200 nM PDBu.

(H) Immunoblot depicting PKC γ WT and P524R phosphorylation. The asterisk denotes phosphorylated and the dash unphosphorylated PKC γ .

(I) Normalized FRET ratio changes (mean \pm SEM) showing PKC activity of PKC β III constructs in COS7 cells co-expressing CKAR. Cells were stimulated with UTP (100 μ M) followed by PDBu (200 nM).

(J) Immunoblot depicting mCherry-tagged PKC β III WT and mutant phosphorylation. The asterisk denotes phosphorylated and the dash unphosphorylated PKC β III.

See also Figure S1.

reverting an endogenous LOF allele to WT would also rescue cell growth. We used DLD1 colon cancer cells because they harbor a PKC β A509T LOF mutation (Figure 3I) to assess whether a heterozygous LOF PKC mutation could confer a survival advantage, as most cancer-associated PKC mutations are heterozygous. We reverted the mutation to WT in three isogenic clones (Figures S2B and S2C) and confirmed that no sequence alterations existed within the top two most likely predicted off-targets (data not shown). Correction of the A509T mutation in the endogenous PKC β (*PRKCB*) allele caused a slight but reproducible increase

in the PKC β levels and a >2-fold increase in PKC α levels, although neither reached statistical significance (Figure 5A). Immunoblot analysis with a phospho-(Ser) PKC substrate antibody revealed significantly higher basal PKC activity in the corrected cells (Figure 5B). This is consistent with the DLD1 parental cells having reduced PKC activity because of the LOF PKC β mutation and the lower PKC α levels. We next tested the ability of these cells to grow in suspension. Consistent with having higher PKC activity and a more tumor-suppressive phenotype, the corrected cells were less viable in suspension (Figure 5C) because they were less capable of forming the compact multicellular aggregates formed by the DLD1 parental cells (Figure 5D). Moreover,

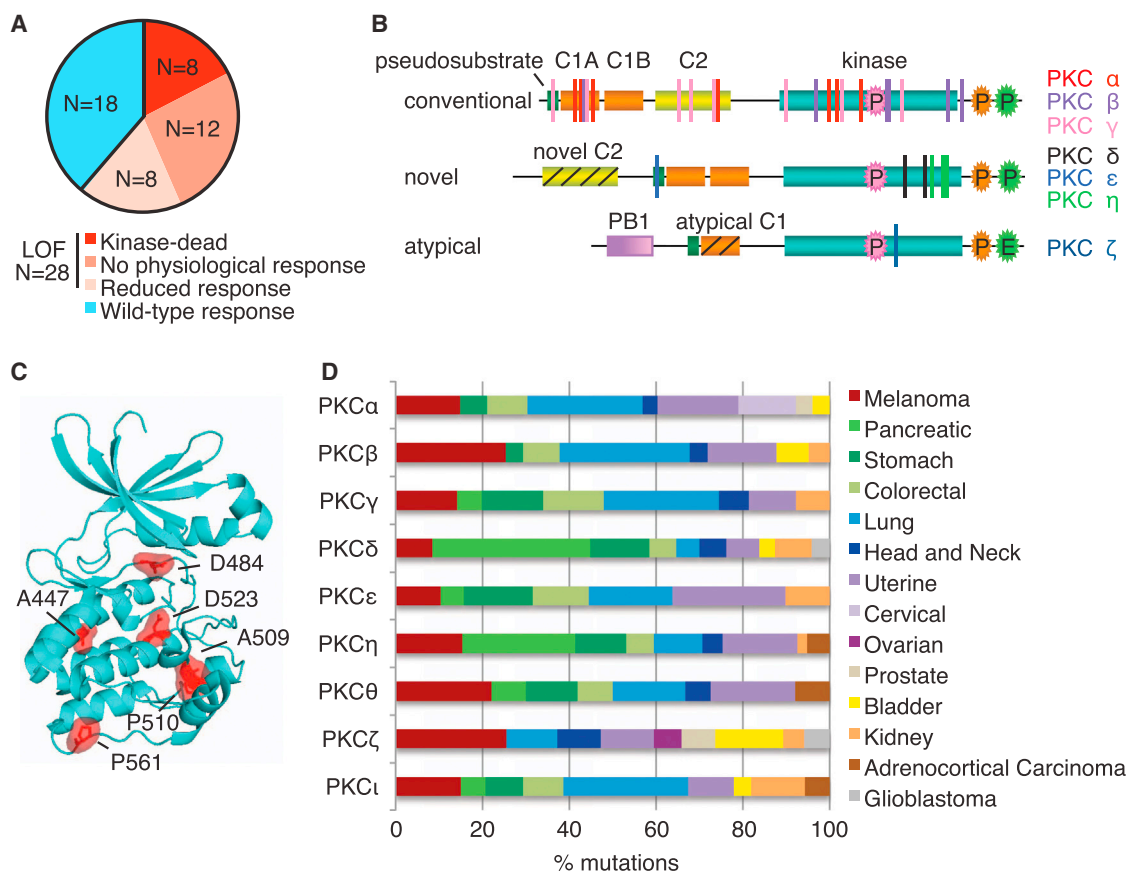


Figure 4. The Majority of PKC Mutations Are LOF

(A) Pie chart of the functional impact of the investigated PKC mutations, with bright red representing mutations that lack any activity, medium red representing mutations that show no response to physiological stimuli (DAG or Ca^{2+} elevation) but some response to non-physiological phorbol esters, light red representing mutations that display reduced activity to physiological stimuli compared to the corresponding WT isozyme, and blue representing no difference from the corresponding WT PKC isozyme.

(B) Domain structure of cPKC, nPKC, and aPKC isoforms, overlaid with the LOF mutations color coded by isozyme.

(C) Crystal structure of the kinase domain of PKC β III (PDB 2I0E) highlighting "warmspot" residues mutated in at least four tumor samples within the various PKC isoforms.

(D) Bar graph depicting the percentage of mutations distributed in the indicated cancers for each PKC isozyme.

the corrected clones had decreased anchorage-independent growth potential (Figure 5E). These results corroborate those obtained from the HCT116 cells overexpressing PKC β II, demonstrating that partial loss of PKC β activity is necessary for growth in soft agar. However, in a 2D proliferation assay, the DLD1-corrected cells proliferated at similar rates to the DLD1 parental cells (Figure S2D), indicating that it is not the proliferation rates that differ between these cells but, rather, their ability to grow in the absence of anchorage.

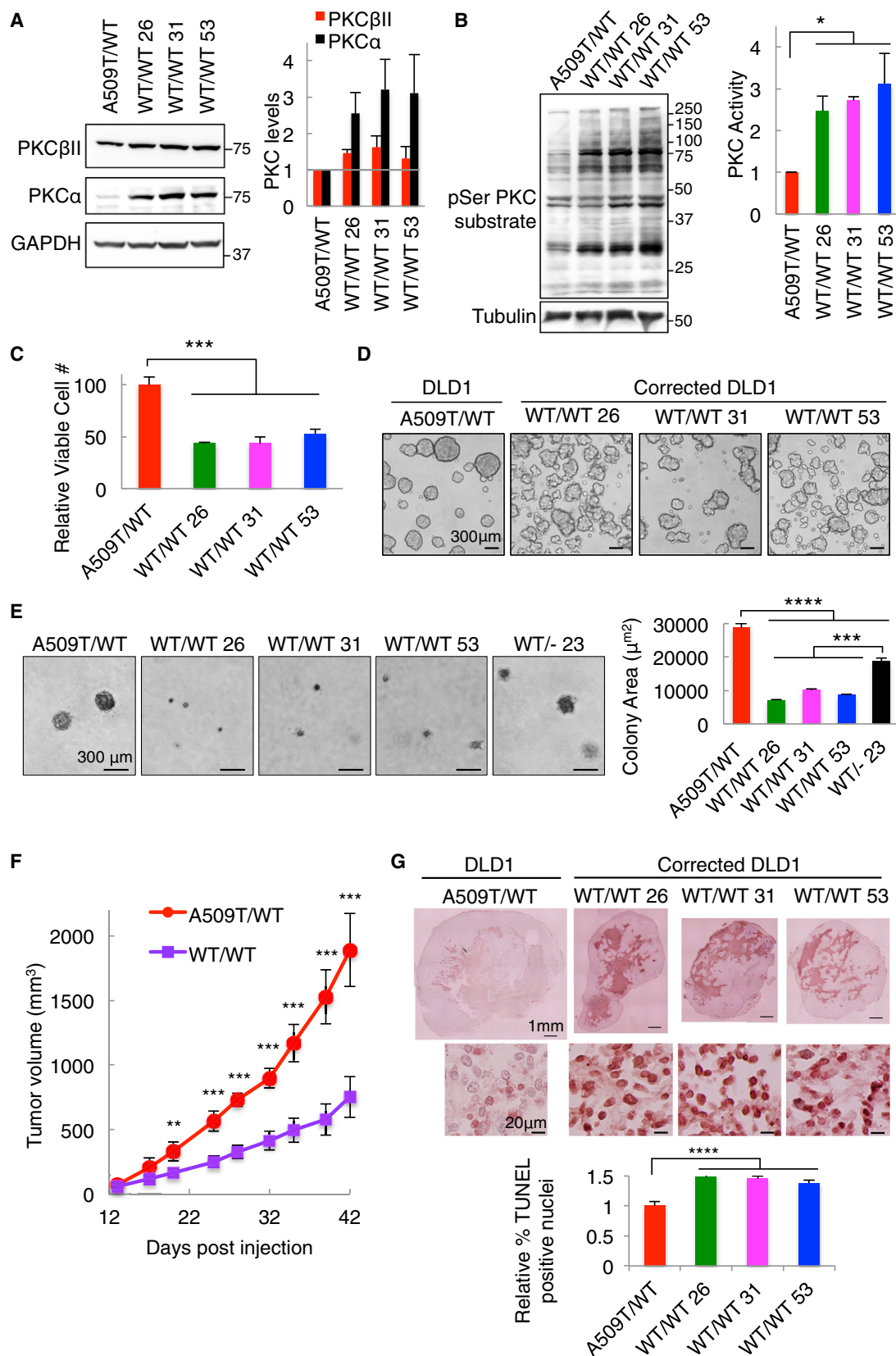
To determine whether PKC displays haploinsufficiency, we knocked out the mutant PKC β allele in DLD1 cells by creating a frameshift deletion using genome engineering (Figure S2E). This hemizygous clone (WT/- 23), containing only one WT allele and thus expressing lower PKC β II levels (Figure S2F), exhibited significantly increased anchorage-independent growth potential compared to cells containing two WT alleles, indicating that PKC β II is haploinsufficient for tumor suppression (Figure 5E). Additionally, the PKC β hemizygous cells did not grow as well

as the PKC β A509T mutated cells in soft agar, indicating that this mutation had a dominant-negative effect.

To definitively establish whether a heterozygous LOF PKC β mutation facilitates tumor growth in vivo, the DLD1 parental or corrected cells were subcutaneously injected into the flanks of nude mice and tumor growth was monitored. Consistent with our cellular data, the tumors derived from the corrected cells were significantly smaller than those from the DLD1 parental cells (Figures 5F and S2G). This reduced growth correlated with increased apoptosis as assessed by TUNEL staining of tumor sections (Figure 5G). These data demonstrate that a heterozygous, dominant-negative PKC β mutation can significantly increase tumor growth, thus establishing PKC β as a tumor suppressor.

DISCUSSION

Here we establish that clinical trials targeting PKC have been based on the wrong assumption; it is not inactivation of PKC



(legend on next page)

but, rather, activation that suppresses tumor growth. Thus, we propose that therapies should target mechanisms to restore the PKC signaling output rather than reduce it. Our comprehensive analysis revealed that 61% of the PKC mutations characterized were LOF and none were activating. We did not account for nonsense mutations or deletions, so an even higher proportion of PKC mutations are LOF. Corroborating our data, three other LOF PKC mutations have been previously described. A LOF PKC α mutation (D294G in C2 domain) was identified in three types of cancer (Alvaro et al., 1993; Prévostel et al., 1997; Zhu et al., 2005) and a LOF PKC ζ mutation (S514F in the kinase domain) was identified in colorectal cancer (Galvez et al., 2009). A partial LOF mutation in PKC ι (R471C), present in three distinct cancers, disrupted substrate binding and induced abnormal epithelial polarity (Linch et al., 2013). To our knowledge, no gain-of-function PKC mutations have been observed in cancer. The identification of LOF mutations throughout the PKC family and in diverse cancers supports a general role for PKC isozymes as tumor suppressors.

Strikingly, several LOF PKC mutations (e.g., PKC β A509V, PKC γ P524R, and PKC α W58L, H75Q, and G257V) acted in a dominant-negative manner by decreasing global endogenous PKC activity. Moreover, the presence of mutant PKC β A509T protein in DLD1 cells reduced PKC α levels. One mechanism for this cross-PKC dominant-negative effect is that the LOF PKC impairs the priming phosphorylations of other PKCs, thus reducing their steady-state levels. This is supported by a prior study demonstrating that unprocessed kinase-dead PKC isozymes prevent the phosphorylation of other PKC isozymes, likely because their phosphorylation requires common titratable components (Garcia-Paramio et al., 1998). This dominant-negative role of LOF mutations is corroborated by studies showing that kinase-dead PKC isozymes function in a dominant-negative manner to exhibit tumorigenic effects on cells (Galvez et al., 2009; Hirai et al., 1994; Kim et al., 2013; Lu et al., 1997). Importantly, although some PKC mutations were dominant negative, loss of PKC such as would occur from nonsense mutations or gene deletions also conferred a growth advantage (Figure 5E), indicating that PKC is haploinsufficient for tumor suppression.

A tumor-suppressive role of PKC is supported by PKC gene knockout mouse models and cellular studies. PKC α -deficient (*Prkca*^{-/-}) mice developed spontaneous intestinal tumors (Oster and Leitges, 2006). In an APC^{Min/+} background, loss of PKC α induced more aggressive tumors and decreased survival (Oster and Leitges, 2006), and in the context of oncogenic Kras, PKC α deletion increased lung tumor formation (Hill et al., 2014). Deletion of PKC ζ in mice that are PTEN haploinsufficient resulted in larger, more invasive prostate tumors and enhanced intestinal tumorigenesis in an APC^{Min/+} background (Ma et al., 2013). Knockdown of PKC δ in colon cancer cells increased tumor growth in nude mice (Hernández-Maqueda et al., 2013). Conversely, overexpression of PKC revealed a protective role. Re-expression of PKC β I in colon cancer cells (Choi et al., 1990) or of PKC δ in keratinocytes (D'Costa et al., 2006) or overexpression of PKC ζ in colon cancer cells (Ma et al., 2013) or in Ras-transformed fibroblasts (Galvez et al., 2009) decreased tumorigenicity in nude mice.

Clinical data reveal lower PKC protein levels and activity in tumor tissue compared with cognate normal tissue, also supporting a tumor-suppressive role for PKC. Total PKC activity was significantly lower in human colorectal cancers versus normal mucosa because of decreased PKC β and PKC δ (Craven and DeRubertis, 1994) or PKC β and PKC ϵ protein levels (Pongracz et al., 1995). PKC α protein was downregulated in 60% of human colorectal cancers (Suga et al., 1998), and PKC ζ was downregulated in renal cell carcinoma (Pu et al., 2012) and non-small cell lung cancer (Galvez et al., 2009). Decreased PKC β and PKC δ levels correlated with increased tumor grade in bladder cancer (Koren et al., 2000; Langzam et al., 2001; Varga et al., 2004), and decreased PKC δ levels correlated with increased grade in endometrial cancer and glioma (Reno et al., 2008; Mandil et al., 2001). PKC η was downregulated in colon and hepatocellular carcinomas, and lower PKC η expression was associated with poorer long-term survival (Davidson et al., 1994; Lu et al., 2009). However, increased PKC ι protein and DNA copy number levels have been observed in certain cancers (Perry et al., 2014; Regala et al., 2005). PKC ι is part of the 3q26 amplicon, and its increased DNA copy number levels correlate with increased mRNA expression (Figure S3). However, DNA copy number

Figure 5. Correction of a Heterozygous LOF PKC β Mutation Reduces Growth in Soft Agar, Suspension, and a Xenograft Model

(A) Immunoblot (left) and quantification (right; mean \pm SEM) of PKC β II, PKC α , and GAPDH levels in the DLD1 cells.

(B) Immunoblot (left) and quantification (right; mean \pm SEM) of phospho-(Ser) PKC substrates. Comparisons were made using a repeated-measures one-way ANOVA followed by post hoc Dunnett's multiple comparison test. * p < 0.05 as compared with the DLD1 parental cells. Data represent the mean of three independent experiments \pm SEM.

(C) Relative viable cell number (mean \pm SEM) as assessed by a trypan blue exclusion assay after 72 hr in suspension from three independent experiments. Comparisons were made by using a one-way ANOVA followed by post hoc Dunnett's Multiple Comparison test. *** p < 0.001 as compared with the DLD1 parental cell group.

(D) Representative phase contrast images of DLD1 cells grown in suspension for 24 hr.

(E) (Left) Colony formation assay in soft agar. (Right) Quantification of colony area (mean \pm SEM) for colonies with a diameter \geq 50 μ m from three to six independent experiments. Comparisons were made using a one-way ANOVA followed by post hoc Tukey's multiple comparison test. **** p < 0.0001 and *** p < 0.001 as compared with the DLD1 parental cell group.

(F) Tumor growth is presented as the mean tumor volume (mm³) \pm SEM, with the red representing data from mice injected with the DLD1 parental cells (A509T/WT; five mice) and purple representing data of the three corrected clones (17 mice total). Comparisons were made using a two-tailed, unpaired Student's t test for each time point. ** p < 0.005 and *** p < 0.0005.

(G) (Top) Representative fields from TUNEL-stained slides of tumors derived from the DLD1 cells. (Bottom) Quantification of TUNEL-positive nuclei (mean \pm SEM). Comparisons were made using a one-way ANOVA followed by post hoc Dunnett's Multiple Comparison test. **** p < 0.0001 as compared with the DLD1 parental cell group.

See also Figure S2.

and mRNA levels do not correlate for cPKC genes (Figure S3). In fact, for PKC α , copy number levels inversely correlate with protein levels in breast cancer (Myhre et al., 2013), the cancer in which PKC α is most amplified (Cerami et al., 2012; Gao et al., 2013). A number of studies reported increased mRNA expression of other PKC genes in cancer; however, mRNA expression and protein levels often poorly correlate (Myhre et al., 2013). Thus, clinical data of this sort are consistent with a tumor-suppressive function of PKC isozymes, although there might be context specific exceptions for PKC ϵ .

The recent discovery that germline LOF mutations in PKC δ are causal drivers of autoimmune lymphoproliferative syndrome and systemic lupus erythematosus, disorders associated with the acquisition of cancer-associated phenotypes, supports a bona fide tumor-suppressive role of PKC in humans (Belot et al., 2013; Kuehn et al., 2013; Salzer et al., 2013). Both diseases are characterized by increased proliferation and decreased apoptosis of B cells (Belot et al., 2013; Kuehn et al., 2013), and patients frequently develop lymphomas (Bernatsky et al., 2005; Mellemkjaer et al., 1997). Moreover, we found that siblings homozygous for a LOF PKC δ mutation have reduced levels of PKC ζ (data not shown), supporting a dominant-negative role of LOF mutations.

How could decreased PKC activity enhance tumorigenesis? One possibility is that PKC isozymes suppress oncogenic signaling by repressing signaling from oncogenes or stabilizing tumor suppressors. Supporting this, unbiased bioinformatic analysis of tumor samples harboring PKC LOF mutations revealed that TP53 (p53) is one of most frequently mutated genes in tumors harboring LOF mutations for each PKC isozyme (Table 2). PKC might promote the tumor-suppressive function of p53 by stabilizing the WT protein. Considerable evidence suggests that phosphorylation by PKC δ stabilizes p53, thus promoting apoptosis (Abbas et al., 2004; Yoshida et al., 2006), but the role of other PKC isozymes is less clear. KRAS was also among the top ten genes mutated in cancers harboring PKC mutations for seven of the PKC isozymes (Table 2), specifically with mutation at Gly12 (Table S3). This argues that PKC might suppress Kras signaling, such that loss of PKC would be required for Kras to exert its full oncogenic potential. Consistent with this, PKC modulates both the activity and localization of Kras through phosphorylation of Ser181 (Bivona et al., 2006). Although the role of this phosphorylation site in tumors remains controversial (Barceló et al., 2014), our analysis is consistent with loss of PKC enhancing its oncogenic potential. In fact, the DLD1 and HCT116 cells used in our assays contained an oncogenic Kras mutation (G13D) that is necessary for the ability of these cells to grow in soft agar (data not shown). This suggests that LOF PKC mutations are not major cancer drivers but, rather, co-drivers that contribute to cancer progression.

We also analyzed which kinase or cancer census genes (genes implicated in cancer) are significantly more commonly mutated (>15-fold) in tumors harboring PKC mutations versus tumors lacking PKC mutations (Table S4). This allowed us to identify proteins that might be important co-drivers or represent novel genetic dependencies for PKC. The tumor suppressor LATS2, which inhibits the Hippo pathway, and the kinases ROCK1 and ROCK2, which are required for the anchorage independent

growth and invasion of non-small cell lung cancer cells, were among the top 20 mutated proteins that were significantly enriched in tumors harboring PKC mutations (Table S4). Our analysis suggests that mutations in these genes provide a greater proliferative advantage upon loss of PKC signaling. We also performed an analysis of cancer-specific genes frequently co-mutated with PKC in lung cancer, colorectal cancer, or melanoma. This revealed very little overlap in co-mutated genes between the three cancers and also between the three classes of PKC isozymes (Table S5), suggesting that the individual PKC isozymes regulate distinct pathways in different cancers. Interestingly, cancers with a high PKC mutation burden, such as melanoma and colorectal cancers, show little PKC amplification. Conversely, cancers that have higher PKC amplification rates, such as breast and ovarian cancers, have few PKC mutations (Cerami et al., 2012; Gao et al., 2013), consistent with PKC mutations having a smaller or different role in breast and ovarian cancers.

The foregoing data provide a mechanism for why inhibiting PKC has proved unsuccessful and, in fact, detrimental in cancer clinical trials: it is not gain of function but, rather, LOF that confers a survival advantage. Therefore, therapeutic strategies should target ways to restore PKC activity. Bryostatin-1, a PKC agonist, also failed as a therapeutic and, in fact, exhibited counter-therapeutic effects in cervical cancer (Nezhat et al., 2004), likely because it downregulates PKC (Szallasi et al., 1994). Therefore, strategies to activate PKC without downregulating it hold significant clinical potential. An important ramification of this study is that drugs that inhibit proteins involved in the processing of PKC cause loss of PKC. Notably, both mTOR and HSP90 inhibitors, currently in use in the clinic (Don and Zheng, 2011; Neckers and Workman, 2012), prevent processing of PKC (Gould et al., 2009; Guertin et al., 2006) and would thus have the detrimental effect of removing its tumor suppressive function. Restoring PKC activity would have to accompany other chemotherapeutics, given that PKC isozymes act as the brakes, not the primary drivers, to oncogenic signaling. Our finding that decreased PKC activity enhances tumor growth challenges the concept of inhibiting PKC isozymes in cancer and underscores the need for therapies that restore or stabilize PKC activity in cells.

EXPERIMENTAL PROCEDURES

FRET Imaging and Analysis

Cells were imaged as described previously (Gallegos et al., 2006). For activity measurements, cells were co-transfected with the indicated mCherry-tagged PKC and CKAR or plasma membrane-targeted CKAR, as indicated. For translocation experiments, cells were co-transfected with the indicated YFP-tagged PKC and membrane-targeted CFP.

Generation of CRISPR Cell Lines

The CRISPR/Cas9 genome-editing system was employed to generate DLD1 cell lines in which the PKC β A509T mutation was reverted to WT or knocked out. For the nuclease method, DLD1 cells were transiently transfected with the hSpCas9 vector containing the gRNA PKC β -a, the PAGE-purified 70-mer ssODN (Figure S2B), and pMAX-GFP. For the double nickase method, DLD1 cells were transfected with two hSpCas9n vectors containing either gRNA PKC β -a or PKC β -b, the ssODN, and pMAX-GFP. GFP⁺ cells were sorted 72 hr later. To reduce off-target mutagenesis, one of the clones (WT/WT 53) was made using a double-nicking approach that requires the

Table 2. Top 20 Genes with Mutations that Co-Occur with PKC Mutations

PKC α (50)	PKC β (90)	PKC γ (102)	PKC δ (47)	PKC ϵ (57)	PKC η (51)	PKC θ (81)	PKC ι (48)	PKC ζ (28)
BLID (7)	<u>TP53</u> (42)	<u>TP53</u> (52)	KRAS (13)	GNG4 (5)	SPINK7 (5)	<u>TP53</u> (42)	SPRR2G (6)	TNP1 (3)
<u>TP53</u> (23)	KRTAP6-2 (6)	CDKN2A (17)	<u>TP53</u> (22)	KRAS (11)	RPL39 (3)	CDKN2A (13)	<u>TP53</u> (26)	<u>TP53</u> (15)
KRTAP19-5 (4)	PCP4 (4)	KRAS (16)	CDKN2A (9)	DEFB114 (4)	KRAS (11)	KRAS (14)	CDKN2A (10)	CNPY1 (3)
SPRR2E (4)	KRAS (12)	HTN1 (4)	CD52 (3)	CNPY1 (5)	DEFB114 (4)	SPANXN5 (5)	BANF1 (5)	SPATA8 (3)
REG3A (8)	OR4A15 (21)	SPRR2G (5)	CNPY1 (4)	SVIP (4)	PLN (3)	DEFB110 (4)	LACRT (7)	SPANXN3 (4)
H3F3C (6)	POM121L12 (18)	DEFB115 (6)	SPINK13 (4)	CXCL10 (5)	DEFB115 (5)	KRTAP15-1 (8)	CXCL9 (6)	KRTAP19-5 (2)
MLLT11 (4)	REG1A (10)	DNAJC5B (12)	ATP5E (2)	KRTAP19-3 (4)	LELP1 (5)	DEFB119 (5)	KRAS (9)	VPREB1 (4)
PI3 (5)	NRAS (11)	REG3G (10)	RPL39 (2)	COX7C (3)	DEFB116 (5)	PPIAL4G (9)	RETNLB (5)	GNG4 (2)
SNURF (3)	PLN (3)	SPATA8 (6)	COX7B2 (3)	KRTAP19-8 (3)	KRTAP19-8 (3)	DPPA5 (6)	WFDC10B (4)	ATP6V1G3 (3)
CDKN2A (7)	GNG4 (4)	REG1A (9)	OR4K1 (11)	SPINK7 (4)	IAPP (4)	CRYGB (9)	DEFB110 (3)	CDKN2A (4)
GNG3 (3)	CDKN2A (9)	POM121L12 (16)	FDCSP (3)	<u>TP53</u> (18)	NPS (4)	SPANXN2 (9)	TMSB15B (2)	DEFB119 (2)
DAOA (6)	DEFA4 (5)	TRAT1 (10)	CARTPT (4)	BANF1 (4)	WFDC10B (4)	KRTAP19-3 (4)	GNG7 (3)	LGALS1 (3)
RPL39 (2)	OR2L13 (16)	HIST1H2AA (7)	DUSP22 (7)	TMSB15B (2)	S100A7L2 (5)	DYNLRB2 (6)	CNPY1 (4)	SCGB1D1 (2)
SVIP (3)	LCE1B (6)	SPINK13 (5)	BANF1 (3)	DEFA4 (4)	CNPY1 (4)	SPATA8 (5)	LSM8 (4)	NANOS2 (3)
PLN (2)	SPANXN3 (7)	CCK (6)	DYNLL2 (3)	POM121L12 (12)	<u>TP53</u> (17)	KRTAP19-8 (3)	KRTAP19-5 (3)	CCL17 (2)
FAM19A2 (5)	KRTAP19-3 (4)	OR4K1 (16)	LYRM5 (3)	GYPA (6)	DPPA5 (5)	RIPPLY3 (9)	SPANXN5 (3)	NRAS (4)
CPLX4 (6)	TRAT1 (9)	OR4A5 (16)	ATP6V1G3 (4)	DYNLRB2 (5)	DEFB131 (3)	POM121L12 (14)	CSTL1 (6)	CCL1 (2)
SEC22B (8)	IFNB1 (9)	CCL7 (5)	DEFB128 (3)	HIST1H2BB (5)	SPINK13 (4)	OR4N2 (14)	DEFA4 (4)	PATE4 (2)
CTXN3 (3)	KRTAP19-8 (3)	B2M (6)	MAP1LC3B2 (4)	HIST1H2BI (5)	RPL10L (9)	DEFB115 (4)	SPANXD (4)	POM121L12 (6)
KRTAP19-3 (3)	KRTAP8-1 (3)	PCP4 (3)	GPX5 (7)	FGFR1OP2 (10)	SPRR2A (3)	OTOS (4)	EDDM3A (6)	CRIP1 (2)

Data were normalized based on gene length, and the number of co-occurring cases is listed in parentheses. Two genes are highlighted: TP53 is underlined, and **KRAS** is in bold.

cooperation between two nickase Cas9 enzymes (Ran et al., 2013). CRISPR-targeted clones were expanded and gDNA was extracted using a Quick-gDNA MiniPrep Kit (Zymo Research Corporation) and were screened for the presence of two wild-type alleles by PCR using primers spanning the A509 locus, followed by restriction digest with BtgZI. This restriction site was only present in the WT allele, and correction of the A509T mutation introduced this site into the other allele. The presence a WT allele at both loci was confirmed by Sanger sequencing (Eton Bioscience).

Xenograft Model

Athymic Nude-*Foxn1*^{nu} mice (Harlan) were housed in compliance with the University of California San Diego Institutional Animal Core and Use Committee. 3×10^6 DLD1 cells in 100 μ l PBS were injected subcutaneously into the right flank of each 4-week-old female mouse. Tumor dimensions were recorded twice weekly and tumor volume was calculated as $1/2 \times \text{length} \times \text{width}^2$. Mice were euthanized 43 days after injection, and tumors were excised. One tumor was excluded, as it did not engraft well (DLD1p), and another was excluded, as it was not subcutaneous (WT/WT 31).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, three figures, and five tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.01.001>.

AUTHOR CONTRIBUTIONS

L.L.G., J.B., T.H., and A.C.N. initiated the study. C.E.A. and A.C.N. conceived the experiments and wrote the manuscript. C.E.A. performed the experiments with assistance from E.K. for imaging and immunoblots and from C.Z. for the xenograft model. F.B.F. advised on the use of the xenograft model. A.M.H., C.W., C.J.M., and J.B. performed the bioinformatic analysis. N.L.S. and E.W.T. made the tetracycline-inducible PKC β II HCT116 cells.

ACKNOWLEDGMENTS

We thank the lab for helpful comments, the Moores Cancer Center Histology Core for the TUNEL staining, Meghdad Rahdar for cell sorting, and Jack Dixon for equipment use. This work was supported by NIH GM43154 to A.C.N., NIH NS080939 and the James S. McDonnell Foundation to F.B.F., and NIH CA82683 to T.H. C.E.A. was supported by the UCSD Graduate Training Program in Cellular and Molecular Pharmacology (T32 GM007752) and the NSF Graduate Research Fellowship (DGE1144086). A.M.H., C.W., N.L.S., E.W.T., C.J.M., and J.B. were supported by Cancer Research UK. T.H. is a Frank and Elise Schilling American Cancer Society Professor and holds the Renato Dulbecco Chair in Cancer Research.

Received: August 21, 2014

Revised: November 12, 2014

Accepted: December 24, 2014

Published: January 22, 2015

REFERENCES

Abbas, T., White, D., Hui, L., Yoshida, K., Foster, D.A., and Bargonetti, J. (2004). Inhibition of human p53 basal transcription by down-regulation of protein kinase Cdelta. *J. Biol. Chem.* 279, 9970–9977.

Alvaro, V., Lévy, L., Dubray, C., Roche, A., Peillon, F., Quérat, B., and Joubert, D. (1993). Invasive human pituitary tumors express a point-mutated alpha-protein kinase-C. *J. Clin. Endocrinol. Metab.* 77, 1125–1129.

Antal, C.E., Violin, J.D., Kunkel, M.T., Skovso, S., and Newton, A.C. (2014). Intramolecular conformational changes optimize protein kinase C signaling. *Chem. Biol.* 21, 459–469.

Barceló, C., Paco, N., Morell, M., Alvarez-Moya, B., Bota-Rabassedas, N., Jaumot, M., Vilardell, F., Capella, G., and Agell, N. (2014). Phosphorylation

at Ser-181 of oncogenic KRAS is required for tumor growth. *Cancer Res.* 74, 1190–1199.

Belot, A., Kasher, P.R., Trotter, E.W., Foray, A.P., Debaud, A.L., Rice, G.I., Szykiewicz, M., Zabot, M.T., Rouvet, I., Bhaskar, S.S., et al. (2013). Protein kinase c δ deficiency causes mendelian systemic lupus erythematosus with B cell-defective apoptosis and hyperproliferation. *Arthritis Rheum.* 65, 2161–2171.

Bernatsky, S., Boivin, J.F., Joseph, L., Rajan, R., Zoma, A., Manzi, S., Ginzler, E., Urowitz, M., Gladman, D., Fortin, P.R., et al. (2005). An international cohort study of cancer in systemic lupus erythematosus. *Arthritis Rheum.* 52, 1481–1490.

Bivona, T.G., Quatela, S.E., Bodemann, B.O., Ahearn, I.M., Soskis, M.J., Mor, A., Miura, J., Wiener, H.H., Wright, L., Saba, S.G., et al. (2006). PKC regulates a farnesyl-electrostatic switch on K-Ras that promotes its association with Bcl-XL on mitochondria and induces apoptosis. *Mol. Cell* 21, 481–493.

Blumberg, P.M. (1980). In vitro studies on the mode of action of the phorbol esters, potent tumor promoters: part 1. *Crit. Rev. Toxicol.* 8, 153–197.

Cacace, A.M., Guadagno, S.N., Krauss, R.S., Fabbro, D., and Weinstein, I.B. (1993). The epsilon isoform of protein kinase C is an oncogene when overexpressed in rat fibroblasts. *Oncogene* 8, 2095–2104.

Castagna, M., Takai, Y., Kaibuchi, K., Sano, K., Kikkawa, U., and Nishizuka, Y. (1982). Direct activation of calcium-activated, phospholipid-dependent protein kinase by tumor-promoting phorbol esters. *J. Biol. Chem.* 257, 7847–7851.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2, 401–404.

Choi, P.M., Tchou-Wong, K.M., and Weinstein, I.B. (1990). Overexpression of protein kinase C in HT29 colon cancer cells causes growth inhibition and tumor suppression. *Mol. Cell. Biol.* 10, 4650–4657.

Craven, P.A., and DeRubertis, F.R. (1994). Loss of protein kinase C delta isoform immunoreactivity in human adenocarcinomas. *Dig. Dis. Sci.* 39, 481–489.

D'Costa, A.M., Robinson, J.K., Maududi, T., Chaturvedi, V., Nickoloff, B.J., and Denning, M.F. (2006). The proapoptotic tumor suppressor protein kinase C-delta is lost in human squamous cell carcinomas. *Oncogene* 25, 378–386.

Davidson, L.A., Jiang, Y.H., Derr, J.N., Aukema, H.M., Lupton, J.R., and Chapkin, R.S. (1994). Protein kinase C isoforms in human and rat colonic mucosa. *Arch. Biochem. Biophys.* 312, 547–553.

Dempsey, E.C., Newton, A.C., Mochly-Rosen, D., Fields, A.P., Reyland, M.E., Insel, P.A., and Messing, R.O. (2000). Protein kinase C isozymes and the regulation of diverse cell responses. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 279, L429–L438.

Don, A.S., and Zheng, X.F. (2011). Recent clinical trials of mTOR-targeted cancer therapies. *Rev. Recent Clin. Trials* 6, 24–35.

Dries, D.R., Gallegos, L.L., and Newton, A.C. (2007). A single residue in the C1 domain sensitizes novel protein kinase C isoforms to cellular diacylglycerol production. *J. Biol. Chem.* 282, 826–830.

Gallegos, L.L., Kunkel, M.T., and Newton, A.C. (2006). Targeting protein kinase C activity reporter to discrete intracellular regions reveals spatiotemporal differences in agonist-dependent signaling. *J. Biol. Chem.* 281, 30947–30956.

Galvez, A.S., Duran, A., Linares, J.F., Pathrose, P., Castilla, E.A., Abu-Baker, S., Leitges, M., Diaz-Meco, M.T., and Moscat, J. (2009). Protein kinase C ζ represses the interleukin-6 promoter and impairs tumorigenesis in vivo. *Mol. Cell. Biol.* 29, 104–115.

Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* 6, p11.

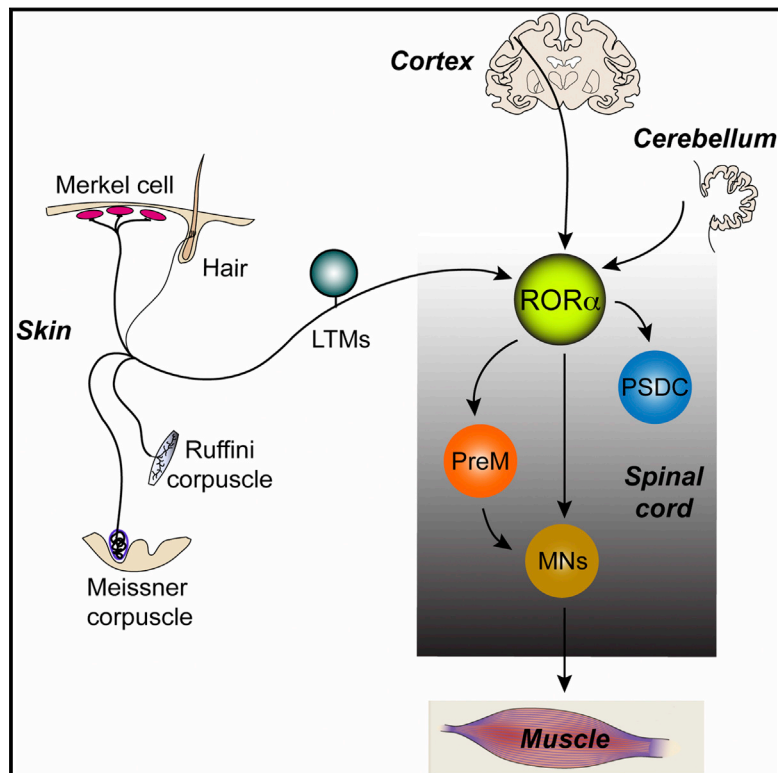
Garcia-Paramio, P., Cabrerizo, Y., Bormancin, F., and Parker, P.J. (1998). The broad specificity of dominant inhibitory protein kinase C mutants infers a common step in phosphorylation. *Biochem. J.* 333, 631–636.

- Gould, C.M., Kannan, N., Taylor, S.S., and Newton, A.C. (2009). The chaperones Hsp90 and Cdc37 mediate the maturation and stabilization of protein kinase C through a conserved PXXP motif in the C-terminal tail. *J. Biol. Chem.* 284, 4921–4935.
- Griner, E.M., and Kazanietz, M.G. (2007). Protein kinase C and other diacylglycerol effectors in cancer. *Nat. Rev. Cancer* 7, 281–294.
- Guertin, D.A., Stevens, D.M., Thoreen, C.C., Burds, A.A., Kalaany, N.Y., Mofat, J., Brown, M., Fitzgerald, K.J., and Sabatini, D.M. (2006). Ablation in mice of the mTORC components raptor, rictor, or mLST8 reveals that mTORC2 is required for signaling to Akt-FOXO and PKC α , but not S6K1. *Dev. Cell* 11, 859–871.
- Gwak, J., Jung, S.J., Kang, D.I., Kim, E.Y., Kim, D.E., Chung, Y.H., Shin, J.G., and Oh, S. (2009). Stimulation of protein kinase C- α suppresses colon cancer cell proliferation by down-regulation of beta-catenin. *J. Cell. Mol. Med.* 13 (8B), 2171–2180.
- Hansra, G., Bornancin, F., Whelan, R., Hemmings, B.A., and Parker, P.J. (1996). 12-O-Tetradecanoylphorbol-13-acetate-induced dephosphorylation of protein kinase C α correlates with the presence of a membrane-associated protein phosphatase 2A heterotrimer. *J. Biol. Chem.* 271, 32785–32788.
- Hernández-Maqueda, J.G., Luna-Ulloa, L.B., Santoyo-Ramos, P., Castañeda-Patlán, M.C., and Robles-Flores, M. (2013). Protein kinase C delta negatively modulates canonical Wnt pathway and cell proliferation in colon tumor cell lines. *PLoS ONE* 8, e58540.
- Hill, K.S., Erdogan, E., Khoo, A., Walsh, M.P., Leitges, M., Murray, N.R., and Fields, A.P. (2014). Protein kinase C ζ suppresses Kras-mediated lung tumor formation through activation of a p38 MAPK-TGF β signaling axis. *Oncogene* 33, 2134–2144.
- Hirai, S., Izumi, Y., Higa, K., Kaibuchi, K., Mizuno, K., Osada, S., Suzuki, K., and Ohno, S. (1994). Ras-dependent signal transduction is indispensable but not sufficient for the activation of AP1/Jun by PKC delta. *EMBO J.* 13, 2331–2340.
- Justilien, V., Walsh, M.P., Ali, S.A., Thompson, E.A., Murray, N.R., and Fields, A.P. (2014). The PRKCI and SOX2 oncogenes are coamplified and cooperate to activate Hedgehog signaling in lung squamous cell carcinoma. *Cancer Cell* 25, 139–151.
- Kang, J.-H. (2014). Protein kinase C (PKC) isozymes and cancer. *New J. Sci.* 2014, 231418.
- Kim, J.Y., Valencia, T., Abu-Baker, S., Linares, J., Lee, S.J., Yajima, T., Chen, J., Eroshkin, A., Castilla, E.A., Brill, L.M., et al. (2013). c-Myc phosphorylation by PKC ζ represses prostate tumorigenesis. *Proc. Natl. Acad. Sci. USA* 110, 6418–6423.
- Koren, R., Langzam, L., Paz, A., Livne, P.M., Gal, R., and Sampson, S.R. (2000). Protein kinase C (PKC) isoenzymes immunohistochemistry in lymph node revealing solution-fixed, paraffin-embedded bladder tumors. *Appl. Immunohistochem. Mol. Morphol.* 8, 166–171.
- Kornev, A.P., Haste, N.M., Taylor, S.S., and Eyck, L.F. (2006). Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proc. Natl. Acad. Sci. USA* 103, 17783–17788.
- Kornev, A.P., Taylor, S.S., and Ten Eyck, L.F. (2008). A helix scaffold for the assembly of active protein kinases. *Proc. Natl. Acad. Sci. USA* 105, 14377–14382.
- Kuehn, H.S., Niemela, J.E., Rangel-Santos, A., Zhang, M., Pittaluga, S., Stoddard, J.L., Hussey, A.A., Evbuomwan, M.O., Priel, D.A., Kuhns, D.B., et al. (2013). Loss-of-function of the protein kinase C δ (PKC δ) causes a B-cell lymphoproliferative syndrome in humans. *Blood* 121, 3117–3125.
- Langzam, L., Koren, R., Gal, R., Kugel, V., Paz, A., Farkas, A., and Sampson, S.R. (2001). Patterns of protein kinase C isoenzyme expression in transitional cell carcinoma of bladder. Relation to degree of malignancy. *Am. J. Clin. Pathol.* 116, 377–385.
- Linch, M., Sanz-Garcia, M., Soriano, E., Zhang, Y., Riou, P., Rosse, C., Cameron, A., Knowles, P., Purkiss, A., Kjaer, S., et al. (2013). A cancer-associated mutation in atypical protein kinase C ι occurs in a substrate-specific recruitment motif. *Sci. Signal.* 6, ra82.
- Lu, Z., Hornia, A., Jiang, Y.W., Zang, Q., Ohno, S., and Foster, D.A. (1997). Tumor promotion by depleting cells of protein kinase C delta. *Mol. Cell. Biol.* 17, 3418–3428.
- Lu, H.C., Chou, F.P., Yeh, K.T., Chang, Y.S., Hsu, N.C., and Chang, J.G. (2009). Analysing the expression of protein kinase C eta in human hepatocellular carcinoma. *Pathology* 41, 626–629.
- Luna-Ulloa, L.B., Hernández-Maqueda, J.G., Santoyo-Ramos, P., Castañeda-Patlán, M.C., and Robles-Flores, M. (2011). Protein kinase C ζ is a positive modulator of canonical Wnt signaling pathway in tumoral colon cell lines. *Carcinogenesis* 32, 1615–1624.
- Ma, L., Tao, Y., Duran, A., Llado, V., Galvez, A., Barger, J.F., Castilla, E.A., Chen, J., Yajima, T., Porollo, A., et al. (2013). Control of nutrient stress-induced metabolic reprogramming by PKC ζ in tumorigenesis. *Cell* 152, 599–611.
- MacKay, H.J., and Twelves, C.J. (2007). Targeting the protein kinase C family: are we there yet? *Nat. Rev. Cancer* 7, 554–562.
- Mandil, R., Ashkenazi, E., Blass, M., Kronfeld, I., Kazimirsky, G., Rosenthal, G., Umansky, F., Lorenzo, P.S., Blumberg, P.M., and Brodie, C. (2001). Protein kinase C α and protein kinase C δ play opposite roles in the proliferation and apoptosis of glioma cells. *Cancer Res.* 61, 4612–4619.
- Mauro, L.V., Grossoni, V.C., Urtreger, A.J., Yang, C., Colombo, L.L., Morandi, A., Pallotta, M.G., Kazanietz, M.G., Bal de Kier Joffé, E.D., and Puricelli, L.L. (2010). PKC Delta (PKC δ) promotes tumoral progression of human ductal pancreatic cancer. *Pancreas* 39, e31–e41.
- Medkova, M., and Cho, W. (1998). Mutagenesis of the C2 domain of protein kinase C- α . Differential roles of Ca $^{2+}$ ligands and membrane binding residues. *J. Biol. Chem.* 273, 17544–17552.
- Mellemkjaer, L., Andersen, V., Linet, M.S., Gridley, G., Hoover, R., and Olsen, J.H. (1997). Non-Hodgkin's lymphoma and other cancers among a cohort of patients with systemic lupus erythematosus. *Arthritis Rheum.* 40, 761–768.
- Mosior, M., and Newton, A.C. (1998). Mechanism of the apparent cooperativity in the interaction of protein kinase C with phosphatidylserine. *Biochemistry* 37, 17271–17279.
- Myhre, S., Lingjærde, O.C., Hennessy, B.T., Aure, M.R., Carey, M.S., Alsner, J., Tramm, T., Overgaard, J., Mills, G.B., Børresen-Dale, A.L., and Sørle, T. (2013). Influence of DNA copy number and mRNA levels on the expression of breast cancer related proteins. *Mol. Oncol.* 7, 704–718.
- Neckers, L., and Workman, P. (2012). Hsp90 molecular chaperone inhibitors: are we there yet? *Clin. Cancer Res.* 18, 64–76.
- Nelson, T.J., and Alkon, D.L. (2009). Neuroprotective versus tumorigenic protein kinase C activators. *Trends Biochem. Sci.* 34, 136–145.
- Newton, A.C. (2003). Regulation of the ABC kinases by phosphorylation: protein kinase C as a paradigm. *Biochem. J.* 370, 361–371.
- Nezhat, F., Wadler, S., Muggia, F., Mandeli, J., Goldberg, G., Rahaman, J., Runowicz, C., Murgo, A.J., and Gardner, G.J. (2004). Phase II trial of the combination of bryostat-1 and cisplatin in advanced or recurrent carcinoma of the cervix: a New York Gynecologic Oncology Group study. *Gynecol. Oncol.* 93, 144–148.
- Oster, H., and Leitges, M. (2006). Protein kinase C alpha but not PKCzeta suppresses intestinal tumor formation in ApcMin/+ mice. *Cancer Res.* 66, 6955–6963.
- Perry, A.S., Furusato, B., Nagle, R.B., and Ghosh, S. (2014). Increased aPKC Expression Correlates with Prostatic Adenocarcinoma Gleason Score and Tumor Stage in the Japanese Population. *Prostate Cancer* 2014, 481697.
- Pongracz, J., Clark, P., Neoptolemos, J.P., and Lord, J.M. (1995). Expression of protein kinase C isoenzymes in colorectal cancer tissue and their differential activation by different bile acids. *Int. J. Cancer* 61, 35–39.
- Prévostel, C., Martin, A., Alvaro, V., Jaffiol, C., and Joubert, D. (1997). Protein kinase C alpha and tumorigenesis of the endocrine gland. *Horm. Res.* 47, 140–144.
- Pu, Y.S., Huang, C.Y., Chen, J.Y., Kang, W.Y., Lin, Y.C., Shiu, Y.S., Chuang, S.J., Yu, H.J., Lai, M.K., Tsai, Y.C., et al. (2012). Down-regulation of PKC ζ in renal cell carcinoma and its clinicopathological implications. *J. Biomed. Sci.* 19, 39.

- Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., and Zhang, F. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154, 1380–1389.
- Regala, R.P., Weems, C., Jamieson, L., Khor, A., Edell, E.S., Lohse, C.M., and Fields, A.P. (2005). Atypical protein kinase C δ is an oncogene in human non-small cell lung cancer. *Cancer Res.* 65, 8905–8911.
- Reno, E.M., Haughian, J.M., Dimitrova, I.K., Jackson, T.A., Shroyer, K.R., and Bradford, A.P. (2008). Analysis of protein kinase C δ (PKC δ) expression in endometrial tumors. *Hum. Pathol.* 39, 21–29.
- Reyland, M.E. (2007). Protein kinase C δ and apoptosis. *Biochem. Soc. Trans.* 35, 1001–1004.
- Rogers, T.B., Inesi, G., Wade, R., and Lederer, W.J. (1995). Use of thapsigargin to study Ca^{2+} homeostasis in cardiac cells. *Biosci. Rep.* 15, 341–349.
- Salzer, E., Santos-Valente, E., Klaver, S., Ban, S.A., Emminger, W., Prengemann, N.K., Garncarz, W., Müllauer, L., Kain, R., Boztug, H., et al. (2013). B-cell deficiency and severe autoimmunity caused by deficiency of protein kinase C δ . *Blood* 121, 3112–3116.
- Suga, K., Sugimoto, I., Ito, H., and Hashimoto, E. (1998). Down-regulation of protein kinase C- α detected in human colorectal cancer. *Biochem. Mol. Biol. Int.* 44, 523–528.
- Sun, S., Schiller, J.H., and Gazdar, A.F. (2007). Lung cancer in never smokers—a different disease. *Nat. Rev. Cancer* 7, 778–790.
- Symonds, J.M., Ohm, A.M., Carter, C.J., Heasley, L.E., Boyle, T.A., Franklin, W.A., and Reyland, M.E. (2011). Protein kinase C δ is a downstream effector of oncogenic K-ras in lung tumors. *Cancer Res.* 71, 2087–2097.
- Szallasi, Z., Smith, C.B., Pettit, G.R., and Blumberg, P.M. (1994). Differential regulation of protein kinase C isozymes by bryostatin 1 and phorbol 12-myristate 13-acetate in NIH 3T3 fibroblasts. *J. Biol. Chem.* 269, 2118–2124.
- Varga, A., Czifra, G., Tállai, B., Németh, T., Kovács, I., Kovács, L., and Biró, T. (2004). Tumor grade-dependent alterations in the protein kinase C isoform pattern in urinary bladder carcinomas. *Eur. Urol.* 46, 462–465.
- Violin, J.D., Zhang, J., Tsien, R.Y., and Newton, A.C. (2003). A genetically encoded fluorescent reporter reveals oscillatory phosphorylation by protein kinase C. *J. Cell Biol.* 161, 899–909.
- Walsh, M.F., Woo, R.K., Gomez, R., and Basson, M.D. (2004). Extracellular pressure stimulates colon cancer cell proliferation via a mechanism requiring PKC and tyrosine kinase signals. *Cell Prolif.* 37, 427–441.
- Wu, B., Zhou, H., Hu, L., Mu, Y., and Wu, Y. (2013). Involvement of PKC α activation in TF/IIa/PAR2-induced proliferation, migration, and survival of colon cancer cell SW620. *Tumour Biol.* 34, 837–846.
- Yoshida, K., Liu, H., and Miki, Y. (2006). Protein kinase C δ regulates Ser46 phosphorylation of p53 tumor suppressor in the apoptotic response to DNA damage. *J. Biol. Chem.* 281, 5734–5740.
- Young, S., Parker, P.J., Ullrich, A., and Stabel, S. (1987). Down-regulation of protein kinase C is due to an increased rate of degradation. *Biochem. J.* 244, 775–779.
- Zhang, L., Huang, J., Yang, N., Liang, S., Barchetti, A., Giannakakis, A., Caudong, M.G., O'Brien-Jenkins, A., Massobrio, M., Roby, K.F., et al. (2006). Integrative genomic analysis of protein kinase C (PKC) family identifies PKC δ as a biomarker and potential oncogene in ovarian carcinoma. *Cancer Res.* 66, 4627–4635.
- Zhang, L.L., Cao, F.F., Wang, Y., Meng, F.L., Zhang, Y., Zhong, D.S., and Zhou, Q.H. (2014). The protein kinase C (PKC) inhibitors combined with chemotherapy in the treatment of advanced non-small cell lung cancer: meta-analysis of randomized controlled trials. *Clin. Trans. Oncol.* Published online October 29, 2014. <http://dx.doi.org/10.1007/s12094-014-1241-3>.
- Zhu, Y., Dong, Q., Tan, B.J., Lim, W.G., Zhou, S., and Duan, W. (2005). The PKC α -D294G mutant found in pituitary and thyroid tumors fails to transduce extracellular signals. *Cancer Res.* 65, 4520–4524.

Identification of a Spinal Circuit for Light Touch and Fine Motor Control

Graphical Abstract



Authors

Steeve Bourane, Katja S. Grossmann, ..., Stephanie Koch, Martyn Goulding

Correspondence

goulding@salk.edu

In Brief

ROR α interneurons in the dorsal spinal cord are essential for processing and transmitting light touch information from innocuous touch receptors in the skin; they also integrate sensory feedback from these touch receptors with motor control information from the cortex and cerebellum to generate the corrective motor movements that animals use when walking on uneven and challenging terrains.

Highlights

- Spinal ROR α interneurons (INs) transmit innocuous light touch information
- ROR α INs are key elements of the spinal touch circuitry for corrective motor movements
- ROR α INs integrate cutaneous sensory feedback with descending motor commands



Identification of a Spinal Circuit for Light Touch and Fine Motor Control

Steeve Bourane,¹ Katja S. Grossmann,¹ Olivier Britz,¹ Antoine Dalet,¹ Marta Garcia Del Barrio,¹ Floor J. Stam,¹ Lidia Garcia-Campmany,¹ Stephanie Koch,¹ and Martyn Goulding^{1,*}

¹Molecular Neurobiology Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA

*Correspondence: goulding@salk.edu

<http://dx.doi.org/10.1016/j.cell.2015.01.011>

SUMMARY

Sensory circuits in the dorsal spinal cord integrate and transmit multiple cutaneous sensory modalities including the sense of light touch. Here, we identify a population of excitatory interneurons (INs) in the dorsal horn that are important for transmitting innocuous light touch sensation. These neurons express the ROR alpha (ROR α) nuclear orphan receptor and are selectively innervated by cutaneous low threshold mechanoreceptors (LTMs). Targeted removal of ROR α INs in the dorsal spinal cord leads to a marked reduction in behavioral responsiveness to light touch without affecting responses to noxious and itch stimuli. ROR α IN-deficient mice also display a selective deficit in corrective foot movements. This phenotype, together with our demonstration that the ROR α INs are innervated by corticospinal and vestibulospinal projection neurons, argues that the ROR α INs direct corrective reflex movements by integrating touch information with descending motor commands from the cortex and cerebellum.

INTRODUCTION

Animals use the sense of touch to identify and discriminate nearby objects, direct motor movements, and reinforce social interactions via affective touch (Abraira and Ginty, 2013; McGlone and Reilly, 2010; Rossignol et al., 2006). Specialized LTMs in the skin detect a variety of touch modalities (Delmas et al., 2011; Li et al., 2011; Vrontou et al., 2013). Skin deformation and vibration are detected by rapidly adapting LTMs that innervate Meissner and Pacinian corpuscles, respectively, while slowly adapting Merkel cells and Ruffini organs respond to indentation and stretch (Delmas et al., 2011; McGlone and Reilly, 2010). C-LTMs and A β /A δ lanceolate LTMs in hairy skin monitor the dynamic and static displacement of hair. Our knowledge of how specific somatosensory modalities, including touch, are relayed and gated within the dorsal spinal cord is more limited, in part because the functional studies undertaken so far have primarily employed genetic mutations that alter a broad swath of neurons (Cheng et al., 2004; Gross et al., 2002; Müller et al., 2002; Wang et al., 2013; Xu et al., 2013). Moreover, it is still unclear how cutaneous touch path-

ways intersect with descending motor pathways to control movement and posture.

The peripheral pathways that transmit thermoceptive, nociceptive, innocuous mechanosensory, and proprioceptive stimuli are highly segregated (Basbaum et al., 2009; Lallmend and Ernfors, 2012). Unmyelinated C-nociceptors and lightly myelinated A δ afferents that relay nociceptive, thermoceptive, and pruritoceptive stimuli project to laminae I/II of the dorsal horn (Lallmend and Ernfors, 2012; Todd, 2010), while large diameter fast conducting neurons that carry proprioceptive information innervate neurons in the intermediate and ventral spinal cord (Jankowska, 1992; Lallmend and Ernfors, 2012). Innocuous touch modalities are transmitted by myelinated A-LTMs and unmyelinated C-LTMs that converge on laminae III-IV (Abraira and Ginty, 2013; Delmas et al., 2011; Lallmend and Ernfors, 2012; Lechner and Lewin, 2013). These and other findings argue that somatosensory information is encoded in the periphery by labeled lines of transmission. However, the extent to which these labeled lines extend into the CNS remains to be determined. It is known that cutaneous LTMs project and arborize within the dorsal horn in a modality-specific manner (Brown, 1981; Fyffe, 1992; Light and Perl, 1979a, 1979b; Shortland and Woolf, 1993; Woolf, 1987). Hair follicle afferents terminate in laminae II and III (Woodbury et al., 2001). Meissner and Merkel cell afferents innervate laminae III and IV (Brown, 1981; Woolf, 1987), while the arbors of Pacinian corpuscle afferents localize to lamina III/dorsal lamina IV and lamina V (Brown, 1981; Semba et al., 1984). Finally, Ruffini organ afferents form collaterals in lamina III and have processes that extend into laminae IV/V (Brown, 1981). This anatomical organization suggests that defined mechanosensory modalities may be transmitted and processed by discrete IN cell types in the spinal cord.

In this study, we show that a class of dorsal spinal cord INs expressing the ROR α nuclear orphan receptor are required for proper light touch perception. Ablating ROR α INs in the caudal spinal cord markedly reduces the behavioral responses mice display to light touch, but not to noxious or itch stimuli. Mice lacking ROR α INs also show a marked increase in foot slips during beam walking, indicating the ROR α INs are necessary for corrective foot movements and fine motor control. This motor deficit, in combination with neuronal tracing experiments showing the ROR α INs are innervated by projection neurons in the lateral vestibular nucleus (LVN) and motor cortex, reveals a role for the ROR α INs in integrating sensory inputs from cutaneous LTMs with descending motor signals from the cortex and vestibular system. In so doing, it provides evidence that

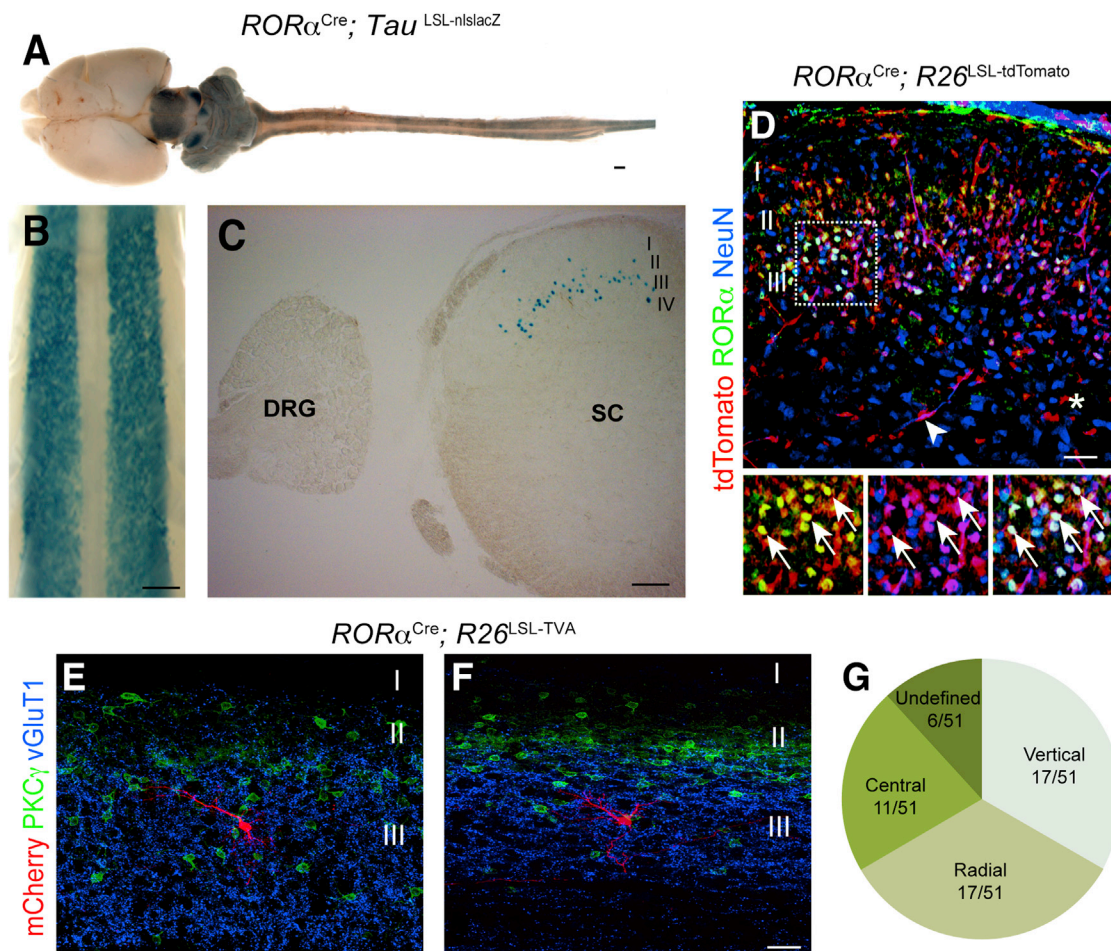


Figure 1. Characterization of $ROR\alpha^{Cre}$ -Derived INs in the Spinal Cord

(A–C) Images from P10 $ROR\alpha^{Cre}; Tau^{LSL-nls lacZ}$ mice showing β -galactosidase expression in the CNS (A) and spinal cord (B and C). (C) shows transverse section through the spinal cord (SC). Note the dorsal root ganglia (DRG) do not express β -galactosidase.

(D) Section through P10 $ROR\alpha^{Cre}; R26^{LSL-tdTomato}$ lumbar spinal cord stained with antibodies to $ROR\alpha$ (green) and NeuN (blue). tdTomato⁺ fluorescence (red) was visualized without staining. Merged image shows tdTomato is largely restricted to $ROR\alpha^{+}/NeuN^{+}$ neurons in laminae Ili/III. The tdTomato⁺ cells located outside of laminae Ili/III do not express NeuN. Arrows indicate double-labeled neurons. Arrowhead in (D) indicates a tdTomato⁺ blood vessel. Asterisk marks a tdTomato⁺ glial cell.

(E–F) mCherry-labeled neurons (red) in the lumbar cord of P21 $ROR\alpha^{Cre}; R26^{LSL-TVA}$ mice. Their location in relation to excitatory $PKC\gamma^{+}$ neurons (green) and vGluT1 sensory afferents (blue) is shown.

(G) Summary of the morphological profile of 51 $ROR\alpha$ INs.

Scale bars, 1,000 μm (A), 500 μm (B), 200 μm (C), 100 μm (D), 25 μm (E and F). See also Figure S1.

the $ROR\alpha$ INs function as an important interneuronal node for coordinating descending motor commands with cutaneous mechanosensory feedback.

RESULTS

$ROR\alpha$ Identifies a Distinct Population of Excitatory Neurons in Laminae Ili/III of the Spinal Cord

The neurons in the spinal cord that express $ROR\alpha$ are primarily restricted to laminae Ili/III (Del Barrio et al., 2013). In mice carrying both the $ROR\alpha^{Cre}$ allele (Chou et al., 2013) and the $Tau^{LSL-nls lacZ}$ reporter allele (this study), β -galactosidase (β -gal) expression was localized to two bilateral columns of neurons in

the dorsal horn (Figures 1A–1C). Most importantly, β -gal was markedly absent from sensory neurons in the dorsal root ganglia (DRG) (Figure 1C). Transverse sections through the spinal cord showed that the $ROR\alpha$ INs are largely restricted to laminae Ili/III, with the intermediate and ventral spinal cord being completely devoid of $ROR\alpha$ expression (Figures 1C and S1A).

To confirm that Cre recombination faithfully recapitulates $ROR\alpha$ expression, P10 $ROR\alpha^{Cre}; R26^{LSL-tdTomato}$ spinal cords were stained with antibodies to DsRed (tdTomato), $ROR\alpha$, and NeuN (Figure 1D). A total of $94.2\% \pm 1.3\%$ of the $ROR\alpha^{+}/NeuN^{+}$ neurons in laminae Ili/III expressed tdTomato. Conversely, $91.8\% \pm 2.9\%$ of the tdTomato⁺/NeuN⁺ neurons expressed $ROR\alpha$. Scattered non-neuronal NeuN-negative tdTomato⁺ cells were present

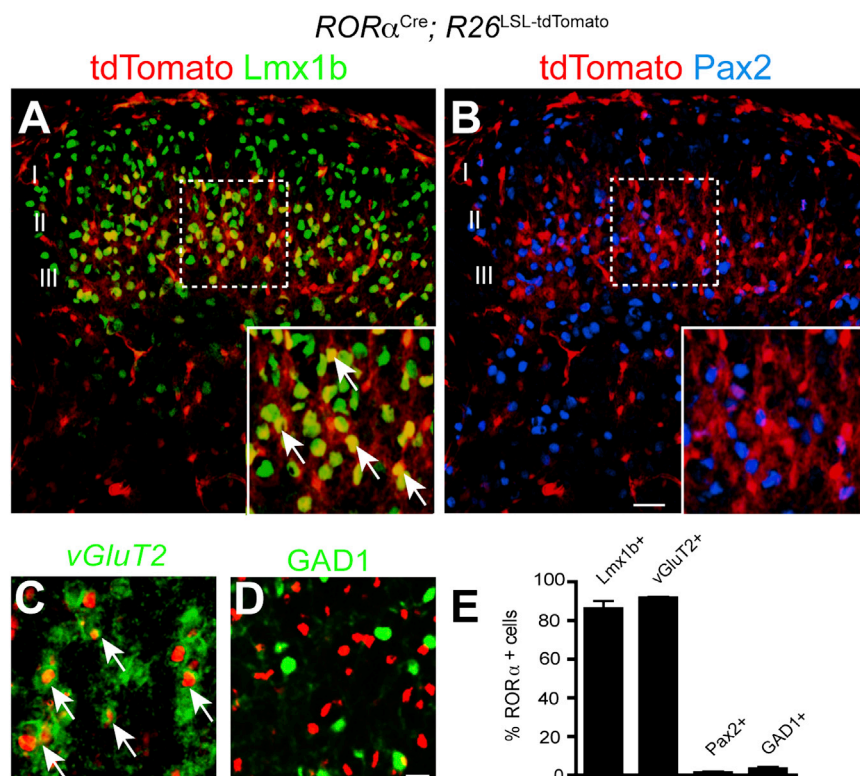


Figure 2. $ROR\alpha$ Identifies a Population of Excitatory Interneurons in the Dorsal Horn of the Spinal Cord

(A and B) Transverse sections through P10 $ROR\alpha^{Cre}; R26^{LSL-tdTomato}$ lumbar dorsal spinal cord showing $ROR\alpha$ is coexpressed with Lmx1b, but not with Pax2.

(C and D) Sections from P10 $ROR\alpha^{Cre}; R26^{LSL-tdTomato}; Gad1-GFP$ spinal cord showing tdTomato is coexpressed with vGluT2 mRNA (C, green), but not GFP (D, green).

(E) Quantification of dorsal spinal cord marker expression in $ROR\alpha$ INs. Cell counts obtained from three spinal cords were pooled and expressed as mean \pm SEM. Cell counts for $ROR\alpha$ INs expressing Lmx1b (940/1093), vGluT2 (904/987), Pax2 (15/1197) and GAD1 (22/845).

Scale bars, 100 μ m (A and B), 10 μ m (C and D). See also Figure S2.

outside of laminae III/III (Figure 1D, asterisk). These cells are predominantly endothelial cells, but also include some glial cells. Anatomical analysis revealed that most $ROR\alpha$ INs possess radial, vertical or central cell-like morphologies (Figure 1G), and have dendritic fields that are largely confined to laminae III/III and dorsal lamina IV (Figures 1E, 1F, S1B and S1C).

In keeping with their morphology, 86.1% \pm 2.7% of the $ROR\alpha$ -tdTomato⁺ INs expressed Lmx1b (Figures 2A and 2E), which marks glutamatergic INs in the dorsal horn (Gross et al., 2002; Müller et al., 2002). By contrast, very few $ROR\alpha$ -tdTomato⁺ INs (1.2% \pm 0.5%) expressed the inhibitory marker Pax2 (Figures 2B and 2E). Further comparison of tdTomato and vesicular glutamate transporter vGluT2 expression confirmed that the $ROR\alpha$ ⁺ INs are primarily glutamatergic, with 91.6% \pm 0.4% coexpressing vGluT2 (Figures 2C, 2E, and S2A–S2C). Moreover, a large fraction of them coexpressed cholecystikinin (Figure S2F), which is restricted to excitatory neurons in the dorsal horn (Hu et al., 2012; Xu et al., 2013). By contrast, only 3.1% \pm 0.9% of the $ROR\alpha$ INs in *Gad1-GFP* mice expressed GFP (Figures 2D and 2E). Further analysis revealed two major molecular subsets of $ROR\alpha$ INs: a dorsal population that colocalizes with PKC γ (Figures S2D and S2F) in lamina II inner, and a more ventral population of MafA⁺/c-Maf⁺/ $ROR\alpha$ ⁺ cells that are restricted to lamina III (Figures S2E and S2F).

Innervation of $ROR\alpha$ INs by Mechanosensory Neurons

In view of their location in laminae III/III, we set out to determine if the $ROR\alpha$ INs receive afferent input from innocuous mechanosensory neurons. CGRP and IB4 stainings were used to compare the location of the $ROR\alpha$ INs with the termination zones of pep-

tidergic and non-peptidergic nociceptive afferents (Todd, 2010). The cell bodies of $ROR\alpha$ INs were located ventral to CGRP⁺ and IB4⁺ afferents in laminae I/II and II, respectively (Figure S3A). Only rarely did we detect putative contacts from nociceptive afferents onto $ROR\alpha$ INs, indicating the $ROR\alpha$ INs receive very little direct nociceptive input. By

contrast, when a c-Ret-CFP transgene (Uesaka et al., 2008), was used to trace the central processes of mechanosensory afferents, we identified numerous CFP⁺/vGluT1⁺ synaptic contacts onto $ROR\alpha$ INs (Figure S3B). These findings provide evidence that the $ROR\alpha$ INs are innervated by cutaneous LTMs as opposed to cutaneous nociceptors.

Monosynaptic retrograde tracing using a pseudotyped EnvA-SADΔG-mCherry rabies virus (Wickersham et al., 2007) was employed to identify the sensory neuron cell types that innervate spinal $ROR\alpha$ INs. These experiments were performed using an intersectional reporter (Li et al., 2013; Stam et al., 2012) that in combination with $ROR\alpha^{Cre}$ and *Cdx2::FlpO* restricts rabies virus entry to $ROR\alpha$ INs in the caudal spinal cord (Figures 3A–3C). Sensory neurons are not directly infected due to their lack of $ROR\alpha$ expression. Injection of EnvA-SADΔG-mCherry rabies virus into the spinal cords of P5 $ROR\alpha^{Cre}; Cdx2::FlpO; R26^{ds-HTB}$ mice, resulted in mCherry expression in multiple cutaneous LTM cell types, including cells with the following expression profiles: c-Ret⁺/IB4[−] A β -LTMs (Figures 3D and 3K; 24.7% \pm 2.3%), TrkC⁺/parvalbumin[−] A β -LTMs (Figures 3E and 3K, 22.0% \pm 2.5%), and TrkB⁺ A δ -LTMs (Figure 3F and 3K, 20.3% \pm 2.3%). Very few IB4⁺ non-peptidergic nociceptors (Figures 3D and 3K, 0.66% \pm 0.33%) and parvalbumin⁺ proprioceptors (Figures 3E and 3K, 3.1% \pm 1.9%) were labeled, demonstrating the $ROR\alpha$ INs are primarily innervated by LTM afferents. We also identified a population of presynaptic neurons that express CGRP and NF200 (Figure 3K and S3G), which are likely to be a subtype of mechanoreceptor (Lawson et al., 2002). Our failure to detect mCherry-labeled tyrosine hydroxylase⁺ C-LTM neurons (Figure 3G; Li et al., 2011) indicates that A β -LTMs and A δ -LTMs,

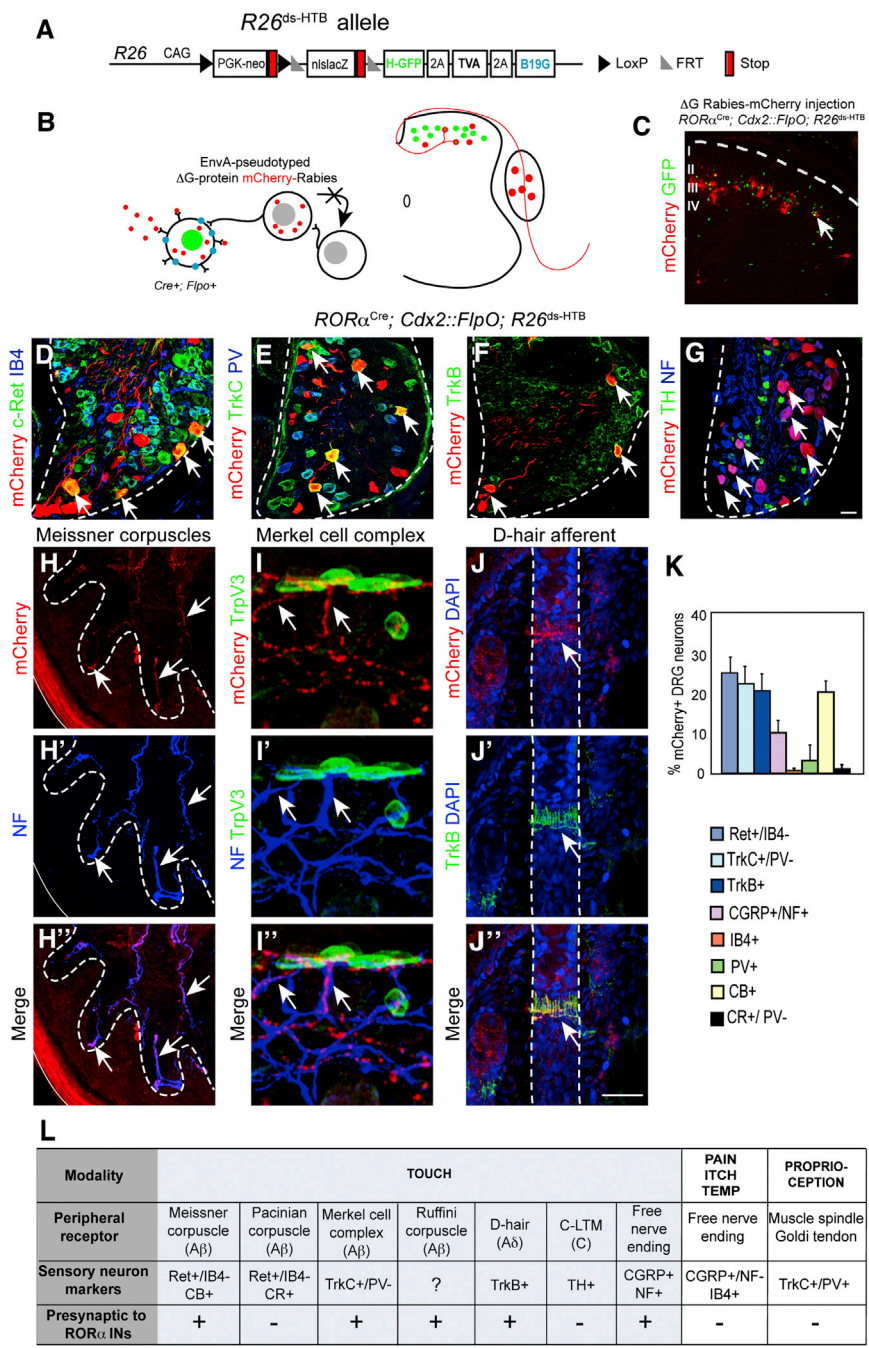


Figure 3. Transsynaptic Rabies Virus Tracing of Sensory Inputs to RORα INs

(A and B) Schematic showing the *R26^{ds-HTB}* allele and strategy used for tracing monosynaptic inputs to the RORα INs using an EnvA-pseudotyped ΔG-mCherry-rabies virus.

(C) Section through P10 *RORα^{Cre}; Cdx2::FlpO; R26^{ds-HTB}* spinal cord showing mCherry-rabies virus labeled cells and RORα INs (green). mCherry⁺/GFP⁺ cells represent transsynaptically labeled presynaptic neurons.

(D–G) Sections from P10 *RORα^{Cre}; Cdx2::FlpO; R26^{ds-HTB}* lumbar DRG showing presynaptically labeled sensory neurons and sensory neuron subtype markers as indicated.

(H) Section of hindlimb footpad stained with antibodies to mCherry (red) and NF (blue) showing Meissner corpuscle innervation.

(I) Section of hindlimb hairy skin stained with antibodies to mCherry (red), NF (blue), and TrpV3 (green), which recognizes Merkel cells.

(J) Section of hairy skin showing a D-hair afferent following staining with antibodies to mCherry (red), TrkB (green) and DAPI (blue).

(K) Quantification of sensory neuronal markers in relation to total number of mCherry⁺ neurons, n = 3 mice.

(L) Summary of sensory afferent types that are presynaptic to the RORα INs.

Abbreviations: NF, neurofilament; CB, calbindin; CR, calretinin; PV, parvalbumin; TH, tyrosine hydroxylase. Dashed lines define the boundary of the spinal cord (C), DRG (D–G), dermal papillae (H) and hair follicles (J). Arrows indicate double-labeled neurons or sensory afferents. Scale bar, 50 μm (D–G), 25 μm (H–J). See also Figure S3.

and their associated Pacinian corpuscle nerve endings did not express mCherry (Figures S3C and S3D), indicating the RORα cells are not innervated by Pacinian LTMs. In hairy skin, we observed mCherry-labeled afferents innervating Merkel cells (Figure 3I). Aδ D-hair terminals (Figure 3J) and transverse lanceolate endings (Figure S3H) were also retrogradely labeled. The morphologies of the hairy skin LTMs labeled by rabies virus are consistent with the expression profile of mCherry in DRG neurons, as D-hair and lanceolate neurons express

rather than C-LTMs, are the primary source of mechanosensory input to the RORα INs.

Our observation that rabies-derived mCherry labels the distal processes of sensory neurons allowed us to unambiguously identify the mechanosensory neuron cell types that innervate the RORα INs. mCherry⁺/NF200⁺ Aβ afferent fibers were present in the dermal papillae of glabrous skin where Meissner corpuscles are located (Figure 3H). This is consistent with mCherry-rabies labeling of calbindin⁺ DRG afferents that innervate Meissner corpuscles (Figure S3E). Calretinin⁺ DRG neurons

a combination of c-Ret, TrkB, or TrkC (Figure 3L) (Lallemend and Ernfor, 2012). mCherry⁺ afferents were occasionally seen together with Ruffini organ endings in the foot muscles (Figure S3F), although the exact extent of RORα IN innervation by Ruffini LTMs remains to be determined.

RORα INs Are Innervated by A-Fiber Afferents and Are Selectively Activated by Light Touch

Whole-cell recordings from RORα INs in lamina III revealed a spectrum of excitatory potentials that closely match the profile of

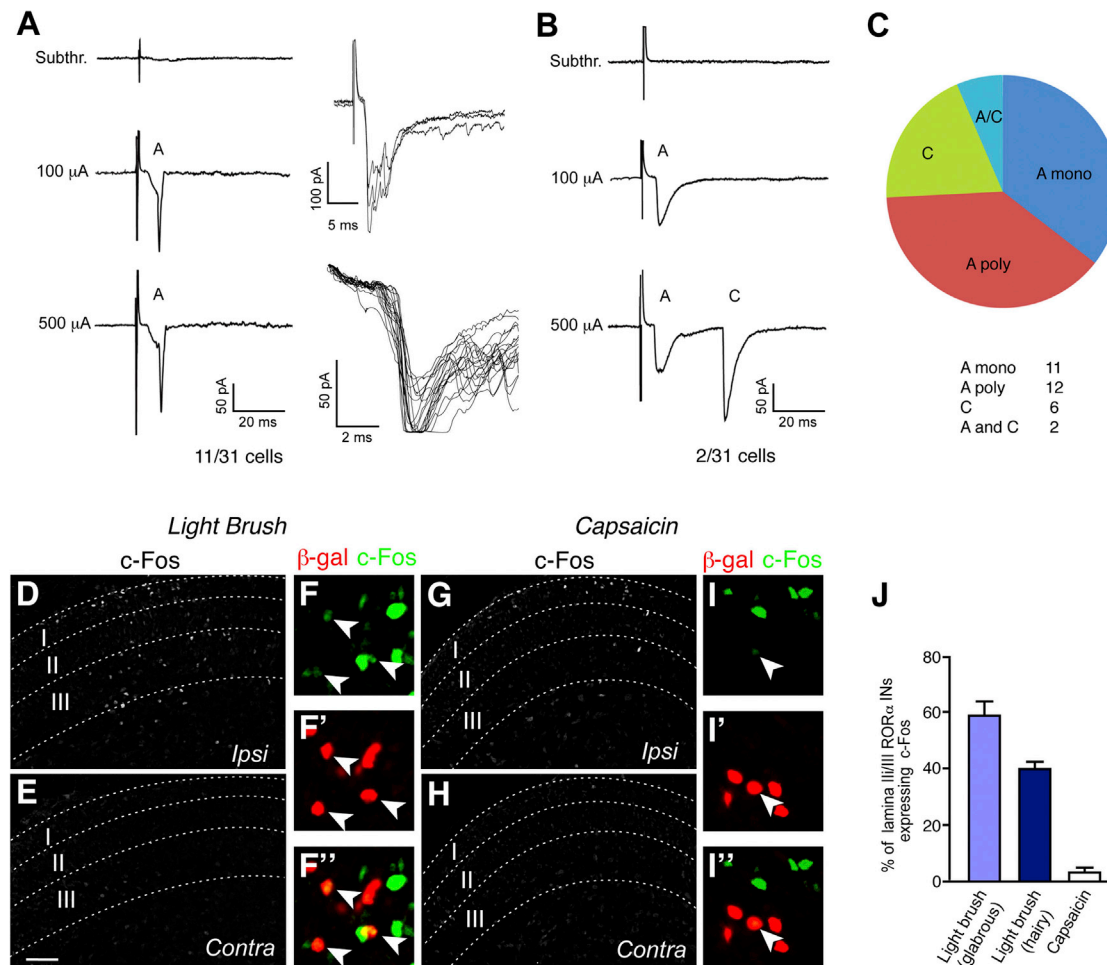


Figure 4. ROR α INs Are Functionally Innervated by Low Threshold Mechanosensory Afferents

(A) Example of a ROR α IN in L4 displaying an A fiber monosynaptic current triggered by low intensity stimulation. Monosynaptic connectivity was confirmed by a 2 Hz frequency stimulation.

(B) Example of a ROR α IN with dual A and C inputs.

(C) Classification of potentials in ROR α INs induced by dorsal root stimulation. $n = 31$ cells.

(D–I) Sections through the lumbar spinal cord of P6 ROR α^{Cre} ; R26^{ds-*HTB*} mice stained with antibodies to c-Fos and β -gal after light brushing (D–F) or injecting capsaicin into the footpad (G–I). Arrowheads in (F) and (I) indicate c-Fos⁺ ROR α INs. (E) and (H) show the contralateral side.

(J) Fraction of ROR α INs (β -gal⁺) coexpressing c-Fos. Only laminae II/III cells located within the domain displaying elevated c-Fos expression were counted. Cell counts of ROR α INs exhibiting c-Fos immunoreactivity: brush stimulation of glabrous skin, 1593/2830 ROR α cells; brush stimulation of hairy skin, 834/2103 ROR α cells; capsaicin injection, 20/711 ROR α cells.

$n = 4$ animals for each experiment. Data: mean \pm SEM. Scale bar, 100 μ m (D, E, G, H).

presynaptic connections seen in the rabies tracing experiments. Short latency-low threshold potentials (< 7 ms) with no failures and low jitter were detected in 11 of 31 cells following dorsal root stimulation (Figure 4A). These potentials most likely represent monosynaptic A β fibers, as Ia muscle spindle afferents do not terminate in lamina III (Brown, 1981; Jankowska, 1992). Low to intermediate threshold potentials that are likely to be either monosynaptic A δ or A polysynaptic inputs were detected in another 12 cells (Figure 4C). Six cells displayed long latency-high threshold potentials (> 25 ms) that are characteristic of polysynaptic C-fiber inputs. Interestingly, two of these cells possessed an additional short latency-low threshold input, indicating some ROR α INs receive a combination of myelinated

A and unmyelinated C fiber innervation (Figure 4B). The exact nature of this polysynaptic C fiber innervation remains to be determined, although it may be derived from C-LTMs that sense pleasant touch (Liu et al., 2007; Vrontou et al., 2013).

We then asked if light brush stimulation activates the immediate early gene c-Fos in ROR α INs. Light brushing of the plantar (underside) surface of the foot (Figures 4D–4F and 4J), and the hairy skin of the lower hindlimb (Figure 4J), resulted in expression of c-Fos in laminae II/III ROR α INs. By contrast, injection of capsaicin into the foot resulted in very little c-Fos induction in ROR α INs ($3.1\% \pm 0.9\%$; Figures 4G–4I and 4J). Instead, c-Fos induction was largely restricted to laminae I/II and lamina V. Some activation of c-Fos was also observed in laminae I/II

neurons in response to brushing (Figure 4D). This may reflect sensitization due to the prolonged brush stimulation needed for c-Fos induction in laminae II/III neurons, or it might be due to the activation of nociceptive pathways by innocuous touch at early postnatal stages (Koch et al., 2012).

Ablating ROR α INs in the Dorsal Spinal Cord Impairs Light Touch

To determine the functional contribution ROR α INs make to somatosensation, we employed a recently developed intersectional *Tau^{ds-DTR}* allele (Figure 5A) to inducibly ablate the ROR α INs in the adult spinal cord. When used in combination with *ROR α ^{Cre}* and a *Cdx2::FlpO* transgene, which is only expressed in caudal regions of the mouse (O.B and M.G, unpublished data), we were able to selectively target diphtheria toxin receptor (DTR) expression to ROR α INs in the caudal spinal cord. ROR α -expressing cells in the anterior CNS and in non-neuronal tissues were not affected by diphtheria toxin (DTX) treatment (Figures 5B and 5C). Using a *R26^{LSL-tdTomato}* reporter to trace the ROR α cells, we estimate that >95% of the ROR α INs in the lumbar cord were ablated by DTX treatment (Figures 5D–5F). Cell killing was largely restricted to excitatory Lmx1b⁺ INs in laminae II/III (Figures 5E and 5G), where ROR α and Lmx1b are coexpressed, with the number of inhibitory Pax2⁺ neurons in the dorsal horn remaining unchanged (Figure 5H).

To confirm cell ablation is restricted to ROR α INs, we analyzed the expression of several excitatory dorsal horn markers (Bröhl et al., 2008; Del Barrio et al., 2013; Gross et al., 2002; Müller et al., 2002). As expected, there was a corresponding reduction in the number of cells expressing PKC γ , MafA, c-Maf, and calbindin (Figures S4A–S4D), while excitatory calretinin⁺ INs that do not express ROR α and GAD1⁺ inhibitory INs were unaffected (Figures S4E and S4F). We also verified that ROR α IN ablation does not alter the laminar organization of sensory afferent terminals in dorsal spinal cord. CGRP⁺ and IB4⁺ nociceptive terminals projected normally to laminae I/II (Figures S4G and S4H) and there was no gross change in the central projections of vGluT1⁺ mechanosensory afferents (Figures S4I and S4J). The termination patterns of peripheral nerve endings (Figures S4K–S4L) were also unchanged following ROR α IN-ablation demonstrating that the loss of these cells does not cause any major change in central and peripheral sensory innervation.

ROR α IN-ablated mice were then subjected to a battery of sensory tests. Strikingly, mice lacking spinal ROR α INs displayed a specific sensory impairment in dynamic and static light touch. In particular, the response to light brushing on the plantar surface of the foot was markedly diminished as compared to control littermates (Figure 5I; 30% \pm 7.9% versus 76% \pm 6.5% of trials). Furthermore, the latency to detecting a piece of sticky tape on the plantar surface of the foot was also increased 2.5-fold (Figure 5J; control 78.9 \pm 17.7 s versus ROR α IN-ablated 192.9 \pm 20.4 s), thereby demonstrating a substantial reduction in responsiveness to static touch. We also observed a marked reduction in tactile sensation on the hairy skin of mice lacking ROR α INs (Figure 5K) indicating the ROR α INs contribute to light touch perception in hairy skin. By contrast, ROR α IN-deficient mice displayed no change in their responses to mechanical pain (Figures 5L–5N)

or sensitivity to heat and cold (Figures 5O and 5P). Responses to chemically induced itch were also unchanged (Figures 5Q and 5R). Finally, there was no difference in acute pain sensation following the injection of capsaicin and formalin (Figures 5S and 5T). In summary, depleting ROR α INs from the dorsal spinal cord results in the selective loss of light touch without altering responsiveness to mechanical, thermal, and chemical pain, and to chemical itch.

ROR α IN-Ablated Mice Display Deficits in Corrective Motor Movements

Our ability to genetically manipulate the ROR α IN population led us to examine the contribution that the ROR α INs make to motor control in mice (Figure 6). Gross behavioral analyses failed to uncover any pronounced change in locomotor activity in ROR α IN-ablated mice (Figure S5), with most measures of locomotor function being largely normal. There was no difference between control and ROR α IN-ablated mice in total distance traveled, vertical activity, or total number of jumps performed during open-field testing (Figures S5A–S5C). Muscle strength, as assessed by the hanging wire test, was also normal in the ROR α IN-ablated mice (Figure S5D). There was also no marked change in gross motor coordination as assessed by the accelerating rotarod and limb coordination during treadmill and ladder beam walking (Figures S5E–S5I).

By contrast, when the raised beam test was used to assess fine motor control during locomotion, a significant increase in hindlimb missteps and slips was seen in ROR α IN-ablated mice traversing a 5 mm beam as compared to control mice (Figures 6A–6C (3.73 \pm 0.73 episodes per crossing for ROR α IN-ablated mice versus 1.29 \pm 0.3 episodes per crossing for control mice)). Most notably, this increase in slips/footfalls was restricted to the hindlimbs, which is consistent with the depletion of ROR α INs at hindlimb levels but not at forelimb levels. In sum, our results demonstrate that ROR α IN-mediated feedback to the motor system is essential for corrective movements, while being largely dispensable for gross motor movements.

ROR α INs Form Synaptic Contacts with Motor Neurons and Premotor Neurons

To further probe the nature of the ROR α IN motor phenotype, we asked if the ROR α INs provide excitatory inputs to motor neurons and molecularly identified premotor IN cell types that control locomotion in mice (Arber, 2012; Goulding, 2009; Grillner and Jessell, 2009; Kiehn, 2011). To visualize the terminal processes of ROR α INs, *ROR α ^{Cre}* mice were crossed with conditional *Thy1^{LSL-YFP}* reporter mice (Buffelli et al., 2003). We observed multiple YFP⁺/vGluT2⁺ contacts on motor neurons in the lumbar lateral motor column (Figure 6D). These vGluT2⁺ contacts are likely to be functional synapses as we find mCherry-positive premotor ROR α INs in the lumbar spinal cords of *ROR α ^{Cre}*; *Tau^{LSL-lacZ}* mice following injection of SADΔG-mCherry rabies virus and AAV-G into the gastrocnemius (ankle extensor) and tibialis anterior (ankle flexor) muscles (Figures 6E–6G). We also looked for YFP⁺/vGluT2⁺ contacts on lamina X V0c cholinergic neurons (Stepien et al., 2010; Zagoraoui et al., 2009), which are the source of muscarinic cholinergic inputs to motor neurons (Miles et al., 2007). ROR α IN-derived contacts were observed on the soma of these cells (Figure 6H), suggesting the ROR α INs are a source

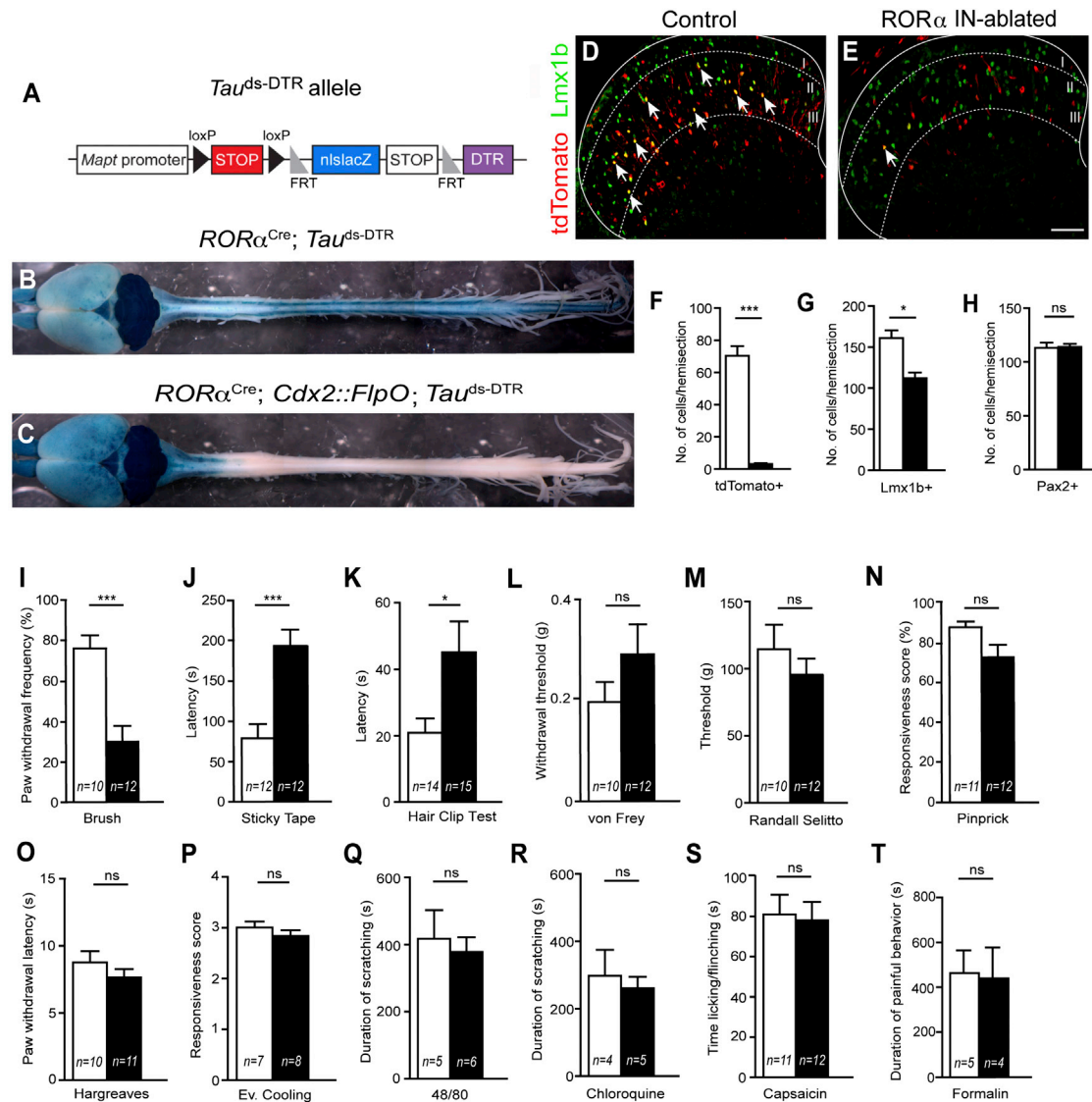


Figure 5. Spinal Ablation of *RORα* INs Impairs Touch but Not Nociceptive Behaviors

(A) Schematic illustrating the *Tau^{ds-DTR}* allele.

(B and C) Brain and spinal cords from P84 *RORα^{Cre}; Tau^{ds-DTR}* (control) and *RORα^{Cre}; Cdx2::FlpO; Tau^{ds-DTR}* (*RORα* IN-ablated) mice showing β-galactosidase reporter expression.

(D and E) Transverse sections through the lumbar dorsal horn of adult control (D) and *RORα* IN-ablated (E) mice 14 days after injecting DTX showing tdTomato (red) and Lmx1b (green) expression. Arrows indicate double-labeled neurons.

(F) *RORα⁺* IN numbers are reduced by 95% following DTX-ablation (3.16 ± 0.46 cells versus 70.30 ± 6.01 cells), *** $p < 0.001$.

(G) Lmx1b-expressing cells are reduced by 30% in *RORα* IN-ablated cords (112.1 ± 6.9 cells versus 161.0 ± 9.5 cells), * $p < 0.05$.

(H) Pax2⁺ cell numbers are unchanged in *RORα* IN-ablated cords (113.1 ± 4.7 cells versus 113.9 ± 2.9 cells), $p > 0.05$.

(I) *RORα* IN-ablated mice show a significant decrease in paw withdrawal to dynamic light brush ($30\% \pm 7.9\%$ versus $76\% \pm 6.5\%$ of trials, *** $p < 0.001$).

(J and K) *RORα* IN-ablated mice show a significant increase in latency to static light touch as measured using the sticky tape test (J, 192.9 ± 20.4 s versus 78.9 ± 17.7 s, *** $p < 0.001$) and detecting a 1.5 mm alligator clip on the hairy skin (K, 47.3 ± 11.5 to 20.9 ± 4.2 , * $p < 0.05$).

(L–N) Mechanical pain measured by von Frey filament, Randall Selitto and pinprick does not differ between control and *RORα* IN-ablated mice.

(O–R) Sensitivity to heat (Hargreaves) and cooling (acetone) in *RORα* IN-ablated mice is normal, as are responses to 48/80 and chloroquine.

(S and T) Chemical pain induced by injecting capsaicin or formalin into the footpad is also unchanged in *RORα* IN-ablated mice.

Data: mean \pm SEM. p values above 0.05 are not significant (ns). n = number of mice tested. Scale Bar: 100 μ m. See also Figure S4.

of excitatory drive to the V0c IN population. DsRed⁺/vGluT2⁺ bouton-like contacts were also found on CFP⁺ V2a INs in *RORα^{Cre}; R26^{LSL}-tdTomato; Chx10^{CFP}* mice indicating putative synaptic connections between *RORα* INs and Chx10⁺ V2a INs

(Figure 6I). These analyses are likely to substantially underestimate the number of putative *RORα* IN-derived synaptic contacts on premotor and motor neurons, given that excitatory contacts are most abundant on dendrites. Taken together, these tracing

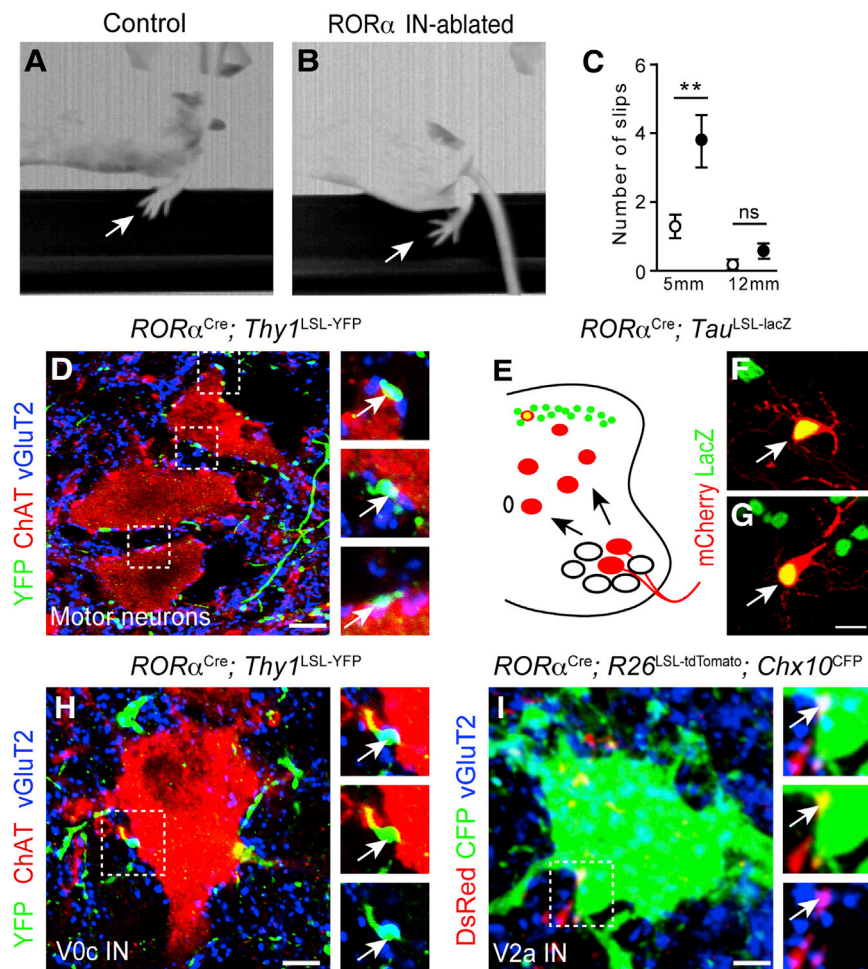


Figure 6. Corrective Movements Are Impaired in RORα IN-Ablated Mice

(A and B) Images showing control (A) and RORα IN-ablated (B) mice crossing a 5 mm wide beam. Arrow indicates position of the foot. Note the position of the foot in B showing a foot slip.

(C) Number of slips and footfalls for 5 mm and 12 mm wide beams. White circles indicate control mice (17 and 11 mice tested for 5 and 12 mm beams, respectively). Black circles indicate RORα IN-ablated mice (19 and 12 mice tested for 5 and 12 mm beams, respectively).

(D and H) P21 RORα^{Cre}; Thy1^{LSL-YFP} spinal cord sections stained with antibodies to GFP (green), vGluT2 (blue) showing synaptic boutons (arrows) on ChAT⁺ motor neurons (D), and ChAT⁺ V0c neurons (H).

(E–G) Retrograde monosynaptic mCherry-rabies virus labeling of RORα INs following injection into the tibialis anterior (F) and gastrocnemius muscles (G). The tracing protocol is shown in (E).

(I) P21 RORα^{Cre}; R26^{LSL-tdTomato}; Chx10^{CFP} spinal cord section stained with antibodies to DsRed (red); GFP (green) and vGluT2 (blue), showing synaptic boutons (arrows) on a Chx10⁺ V2a IN (green).

Data: mean ± SEM, **p < 0.01, ns, p > 0.05. Scale bars, 5 μm (D), 10 μm (F and G), 2 μm (H and I). See also Figure S5.

studies provide evidence that excitatory RORα INs relay touch information from cutaneous LTMs to the spinal motor system.

In view of the contribution that the corticospinal tract (CST) and lateral vestibulospinal tract (LVST) pathways make to fine motor control and balance (Armstrong, 1988; Ito, 2012), we asked whether CST neurons and LVST neurons synapse directly onto the RORα INs in the lumbar spinal cord. Targeted infection of lumbar level RORα INs with EnvA-SADΔG-mCherry rabies virus resulted in monosynaptic rabies virus labeling of LVST neurons in the lateral vestibular nucleus (Figure 7A). These cells, which are innervated by calbindin⁺ Purkinje cells (Figure 7A, inset, arrows), constitute a major vestibular efferent pathway from the cerebellum to the spinal cord (Ito, 2012). We also found multiple mCherry-labeled neurons (Figure 7B) in lamina V of the contralateral mouse primary motor cortex that had the typical pyramidal morphology of corticospinal projection neurons (Figures 7B and 7C). When whole-cell recordings were performed on RORα INs in the lumbar cord (L4), eight of ten RORα INs displayed monosynaptic excitatory potentials in response to stimulating A fiber sensory afferents and the ventral dorsal funiculus that contains the axons of corticospinal projection neurons (Figures 7D–7F). The protocol that we used has been shown to selectively activate descending corticospinal axons (Hantman and Jessell, 2010),

RORα INs are decorated with PKCγ⁺/vGluT1⁺ and Emx1-GFP⁺/vGluT1⁺ contacts from corticospinal projection neurons (Figures 7G and 7H). These same cells are also contacted by vGluT1⁺ processes, which are likely to be derived from mechanosensory afferents (Figures 7G and H, arrowheads).

DISCUSSION

This study identifies a specific class of neurons in the dorsal spinal cord that has an essential role in sensing light touch. The RORα INs transmit innocuous mechanical stimuli from both the hairy and glabrous skin, and depleting the spinal cord of RORα INs leads to a selective mechanosensory deficit that closely matches the repertoire of inputs these cells receive from cutaneous LTMs. We propose that the RORα INs serve as an integrative node that merges sensory input from cutaneous LTMs with descending signals from the cortex and cerebellum to generate the postural adjustments and corrective foot movements that are used to counteract foot slippage.

Coding of Mechanical Stimuli by RORα INs

Our results showing discriminative light touch behaviors are impaired in RORα IN-ablated mice (Figures 5I and 5J), whereas

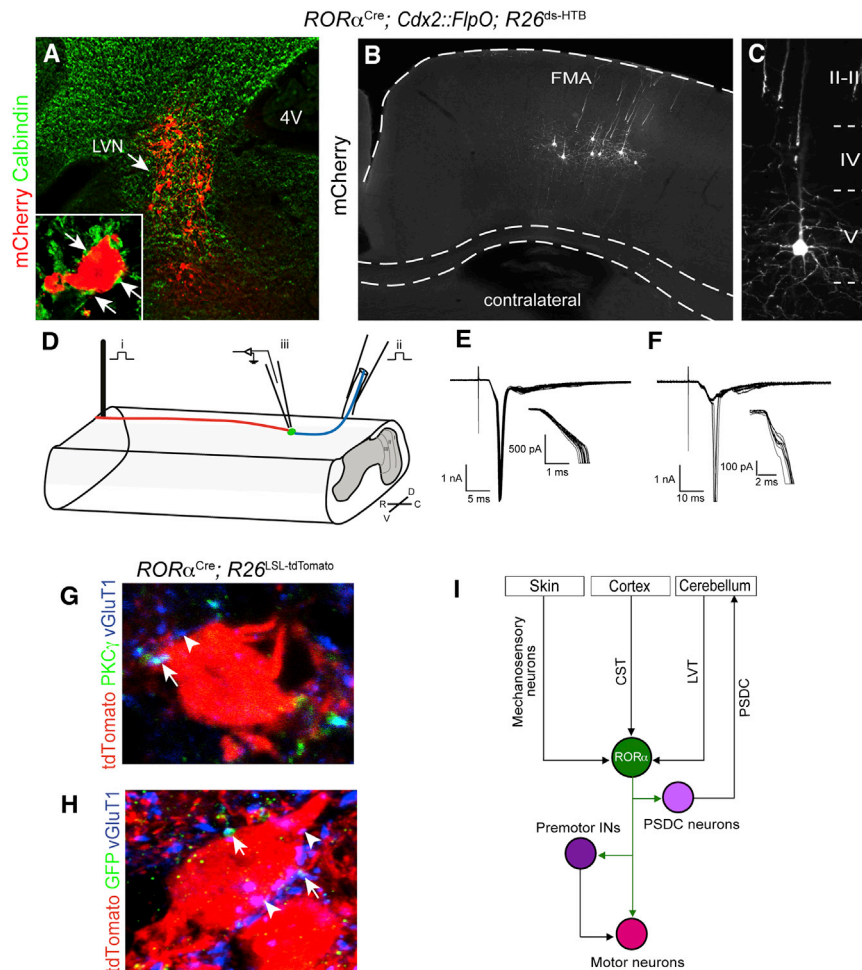


Figure 7. Convergence of Sensory and Descending Inputs onto RORα INs

(A–C) Transverse sections from a P15 *RORα^{Cre}; Cdx2::FlpO; R26^{ds-HTB}* brain 5 days after infecting RORα INs in the lumbar cord with an EnvA G-deleted rabies-mCherry virus. Transsynaptically labeled neurons (red) in the lateral vestibular nucleus (LVN) are innervated by calbindin⁺ Purkinje cells afferents (green, A). Pyramidal neurons in the frontal motor cortex area (FMA) are transsynaptically labeled (B and C).

(D) Hemicord preparation showing the stimulation-whole patch recording set-up for RORα INs. The region of the dorsal funiculus that contains corticospinal axons was stimulated with a concentric bipolar electrode 5 segments (3–4 mm) rostral to the recorded cell (i). A single dorsal root (L4–L6) was stimulated with a glass suction electrode (ii). RORα cells (green) in L4 were patched and EPSCs were recorded (iii). R, rostral; C, caudal; D, dorsal; V, ventral.

(E) Representative recording of an A fiber-type dorsal root evoked potential in a RORα IN. Magnification of onset is shown.

(F) Putative corticospinal tract evoked EPSCs in the RORα IN that is shown in (E). Latency and jitter properties are consistent with monosynaptic connectivity (see magnification of onset).

(G) Section from a P21 *RORα^{Cre}; R26^{LSL-tdTomato}* spinal cord showing PKCγ⁺/vGluT1⁺ corticospinal processes contacting a RORα IN (red). Arrow indicates a PKCγ⁺/vGluT1⁺ corticospinal-derived synaptic bouton. Arrowhead indicates a vGluT1⁺ sensory contact.

(H) Section from a P21 *RORα^{Cre}; R26^{LSL-tdTomato}; Emx1^{GFP}* spinal cord stained with antibodies to DsRed (red), GFP (green) and vGluT1 (blue), which

confirms GFP⁺/vGluT1⁺ corticospinal contacts (arrows) onto RORα INs (red). Arrowhead indicates a vGluT1⁺ sensory contact.

(I) Schematic showing the RORα IN connectivity and an outline of the RORα IN spinal circuit for light touch. NB. Arrows to and from the cerebellum do not indicate direct connections.

Scale bar, 2 μm (G and H). See also Figure S6.

nociceptive behaviors are not (Figure 5), provides evidence that somatosensory inputs to the spinal cord are processed and transmitted in a modality-specific manner. This finding, together with studies showing excitatory neurons located dorsal to the RORα INs in lamina II transmit mechanical pain (Duan et al., 2014), is consistent with the existence of ‘labeled’ interneuronal lines of transmission within the CNS, and it supports a model in which innocuous versus noxious touch modalities are routed through different interneuronal pathways in the spinal cord. We propose that light touch is gated by RORα INs in laminae II/III, whereas mechanical pain is processed by somatostatin⁺ neurons in lamina II, which is in accordance with the general organization of nociceptive and mechanosensory projections in the dorsal spinal cord (Abraira and Ginty, 2013; Lechner and Lewin, 2013; Todd, 2010).

The RORα INs are innervated by multiple LTM subtypes (Figures 3 and S3), indicating that they process multiple streams of mechanosensory input from the skin. In the trigeminal vibrissa

system, tactile inputs converge on projection neurons that receive tactile information from slowly adapting Merkel cell afferents and from rapidly adapting lanceolate afferents (Sakurai et al., 2013). The RORα INs differ from these projection neurons in that they are exclusively local circuit neurons (Figures 1 and S1). The majority of RORα INs appear to be innervated by a single sensory afferent fiber type, with 29 of 31 recorded cells displaying a single depolarizing potential following dorsal root stimulation (Figure 4). Although this finding is consistent with innervation by a single mechanosensory cell type, we cannot exclude the possibility that individual RORα INs receive more than one type of mechanosensory input. For example, dual depolarizing potentials consistent with monosynaptic A and polysynaptic C inputs were detected in two RORα cells (Figure 4B), raising the possibility that some RORα cells merge sensory information from A-LTMs and C-LTMs.

Our observation that the RORα INs are anatomically and molecularly heterogeneous (Figures S1 and S2), suggests that

subpopulations of ROR α INs may process different types of mechanosensory information. Different LTM cell types are known to innervate discrete laminar territories in the dorsal horn (Li et al., 2011; Schouenborg, 2008), raising the possibility that different forms of touch, e.g., caress or pressure, activate different combinations of ROR α IN subtypes. For instance, PKC γ^+ /ROR α^+ INs in lamina II are likely to be innervated by A δ afferents (Li et al., 2011; Light and Perl, 1979a, b; Todd, 2010), while ROR α^+ /MafA $^+$ /c-Maf $^+$ neurons in lamina III would be activated by A β -LTM inputs. Sensory coding by the ROR α INs may therefore play an important role in discriminating different forms of touch.

One question that arises from this study is whether innocuous touch is mediated solely by the ROR α INs or whether other excitatory IN cell types in the dorsal horn also contribute to the sense of light touch. The deficits in sensing light touch the ROR α IN-ablated mice are very pronounced in the glabrous skin, indicating that light touch information from the glabrous skin is primarily gated via ROR α INs. However, it is likely that other excitatory INs contribute to innocuous touch transmission from hairy skin, as ROR α INs are not directly innervated by TH $^+$ C-LTMs (Figure 3). ROR α INs also lack inputs from Pacinian corpuscles (Figure S3), with the MafA $^+$ /c-Maf $^+$ neurons in lamina IV that lack ROR α expression being the most likely candidates for processing information from Pacinian LTMs.

ROR α INs: A Nexus for Integrating Cutaneous Touch and Descending Motor Control Signals

Human studies have shown that the cutaneous sensory, visual, and vestibular systems all contribute to fine motor control and balance (Perry et al., 2000; Stål et al., 2003). It has also been shown in cats, that the corticospinal pathway displays functional convergence on cutaneous reflex pathways (Bretzner and Drew, 2005; Lundberg and Voorhoeve, 1962). What was not clear was how information from these different pathways is merged to provide a coherent set of commands to the spinal motor system. Our discovery that ROR α INs are innervated by descending motor pathways from the cortex and cerebellum suggest that much of this integration may occur at the level of the dorsal horn with the ROR α INs playing a prominent role in integrating cutaneous sensory information with descending motor commands. We propose that the ROR α INs, via their direct excitatory inputs to premotor neurons and motor neurons, function as the core integrative element of a sensorimotor circuit for corrective motor behaviors and fine motor control (Figure 7I).

Mechanosensory feedback from the sole of the foot makes a major contribution to postural stability (Perry et al., 2000; Stål et al., 2003), with these pathways often being compromised in elderly and Parkinson's patients that are prone to falling (Patel et al., 2009; Prätorius et al., 2003; Zia et al., 2003). With respect to the motor phenotype in mice lacking ROR α INs, the loss of sensory feedback from Meissner corpuscles in the sole of the foot that are used to sense slippage would contribute to the increase in foot slips, and possible issues with balance, during beam walking. Deficits in detecting skin deformation and edges due to impaired signaling from Merkel cells and Ruffini endings may also be a factor.

The ROR α INs are innervated by LVST neurons, which points to an important role for the ROR α INs in gating the output of

LVN to the spinal cord. LVST neurons functionally facilitate limb extension (Grillner et al., 1971; Hultborn et al., 1976), in part via their excitatory actions on Ia reciprocal inhibitory neurons. Our observation that a subpopulation of V2b INs possess the features of Ia reciprocal inhibitory neurons (Zhang et al., 2014), coupled with rabies-tracing experiments showing the ROR α INs synapse onto V2b INs (FS and MG; unpublished data), raises the possibility that the vestibular control of limb extension may be partly mediated by the excitatory actions of ROR α INs on V2b-derived Ia reciprocal inhibitory neurons that inhibit flexor motor neurons.

ROR α IN-dependent sensory feedback is largely dispensable for gross locomotor movements. This finding concurs with studies in the cat showing gross locomotor behaviors such as walking are largely independent of cutaneous feedback (Rossignol et al., 2006). Nonetheless, light touch can exert strong phase-dependent effects on locomotion, as exemplified by the stumbling corrective response and paw shake reflex (Forssberg, 1979; Quevedo et al., 2005; Rossignol et al., 2006), with the ROR α INs being potential candidates for mediating these corrective reflexes. The limited role that light touch plays in shaping coarse stepping movements suggests the sensorimotor system is functionally organized so that light touch is primarily used for corrective motor movements. In this way, vestibulospinal, rubrospinal, and corticospinal pathways that intersect with cutaneous sensory pathways in the dorsal spinal cord would have the capacity to elicit dynamic changes in posture and limb/foot/digit position without disrupting the overall locomotor pattern. Under certain conditions, however, these pathways are able to alter the locomotor pattern.

Sensorimotor circuits in the spinal cord and cerebellum are organized into action-based sensorimotor modules that constitute a functional scaffold for corrective and/or reflexive motor behaviors (Schouenborg, 2008). We propose that the ROR α IN touch circuit represents an action-based sensorimotor circuit for corrective motor behaviors. Specifically, the ROR α INs couple a tactile map of the body surface to motor pathways via their connections to elements of the spinal motor circuitry (Figure 6) and to postsynaptic dorsal column (PSDC) projection neurons (Figure S6) that relay tactile sensory information to the cerebellum (Ito, 2012). In this context, the ROR α IN-PSDC-cerebellar pathway would operate as an error detection system to correct motor movements in response to tactile sensory feedback (Apps and Garwicz, 2005). In summary, the ROR α IN-PSDC-cerebellum pathway and the reciprocal LVST-ROR α IN pathway constitute a spinocerebellar feedback loop that utilizes cutaneous sensorimotor feedback to shape the fine corrective movements animals use for dynamic motor control.

EXPERIMENTAL PROCEDURES

Mouse Lines

Multiple mouse lines were used for this study. Mice were maintained on a mixed background and littermates were used as controls for all experiments. ROR α^{Cre} mice were kindly provided by Dr. Dennis O'Leary (Chou et al., 2013). Littermates lacking the *Cdx2::FlpO* allele were used as controls for all behavioral experiments. Expression analyses were performed using the R26^{LSL-tdTomato} (Madisen et al., 2010), *Tau*^{LSL-nlslacZ} (FRT-stop-FRT deleted *Tau*^{ds-DTR}; Duan et al. 2014), and R26^{ds-HTB} mice (Stam et al., 2012). R26^{ds-HTB} mice R26^{LSL-TVA} mice (Seidler et al., 2008) were used for

transsynaptic tracing and morphological analysis with EnvA-pseudotyped rabies virus. *Tau^{ds-DTR}* and *Cdx2::FlpO* mouse lines were used for all the behavioral analyses.

Immunohistochemistry

Immunohistochemical analyses were performed on cryostat sections of fixed tissues using previously described methods (Bourane et al., 2007; Gross et al., 2002).

In Situ Hybridization

In situ hybridization was performed as previously described (Bourane et al., 2007).

c-Fos Induction

P6 *ROR α ^{Cre}*; *R26^{ds-HTB}* mice, restrained in a specially designed enclosure, were acclimated for 10 min prior to stimulation. For light brush stimulation, a soft paint brush was used to gently stroke the plantar surface (glabrous) or dorsal surface (hairy) of the hindpaw for 45 min at approximately 0.5 Hz. To activate pain pathways, 6 μ l of capsaicin (1 mg in 10 ml saline, 7% Tween-80) was injected subcutaneously into the plantar surface of the hindpaw. A 60 min chase time was included between stimulation and sacrificing the animals. Spinal cords were immediately dissected and fixed in 4% paraformaldehyde/PBS. Frozen sections from the lumbar spinal cord were immunostained with antibodies specific for c-Fos and β -galactosidase.

Rabies Virus Tracing

Injections of pseudotyped rabies virus into the lumbar spinal cord of P14 *ROR α ^{Cre}*; *R26^{LSL-TVA}* mice were used to examine ROR α IN morphology. For transsynaptic tracing studies, injections were made into the lumbar cord of *ROR α ^{Cre}*; *Cdx2::FlpO*; *R26^{ds-HTB}* animals at P5 or P10. For muscle injections, single muscles were injected with G-protein-deleted-mCherry rabies virus according to (Stepien et al., 2010).

Behavioral Testing and Analysis

All the behavioral tests were performed blind to the genotype of the animals. Animal experiments were conducted according to NIH guidelines using protocols approved by the Salk Institute for Biological Studies IACUC. Detailed protocols for all behavioral tests are described in [Extended Experimental Procedures](#).

Electrophysiology

Whole-cell recordings and dorsal root stimulation were performed using sagittal hemisections prepared from postnatal P5–P21 *ROR α ^{Cre}*; *Thy1^{LSL-YFP}* mice as described by (Torsney and MacDermott, 2006), with minor modifications. The CST was stimulated with a concentric bipolar electrode (FHC) positioned 3–4 mm rostral to the recorded neuron. Stimulation was performed at 2 Hz (100 μ A, 0.1 ms) to exclude retrograde axonal transmission. See [Extended Experimental Procedures](#).

Quantitative Analysis and Statistics

Cell counts were determined by analyzing 3–6 spinal cords (5–10 sections each) per genotype. All data are presented as the mean \pm SEM with n indicating the number of mice analyzed. Statistical analyses were performed by two-tailed, unpaired Student's t test. p values below 0.05 were considered to be statistically significant.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures and six figures and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.01.011>.

AUTHOR CONTRIBUTIONS

S.B., K.S.G., S.K., M.G.D.B., and L.G.-C. characterized the ROR α INs and performed the behavioral analyses. A.D. performed the electrophysiological

recordings. O.B., and F.J.S. generated the mice and contributed to the experimental analysis. S.B. and M.G. wrote the manuscript. M.G. designed and supervised the study.

ACKNOWLEDGMENTS

We thank Chris Kintner, Qiufu Ma, and John Thomas for discussions and helpful comments on this study. We acknowledge Bo Duan for advice on behavioral tests and Catherine Farrokhi for help with the rotarod and open field analysis. This work was supported by the NIH (NS080586, NS086372, NS072031) to M.G., the Catharina Foundation to S.B., and the Humboldt Foundation to K.S.G.

Received: August 25, 2014

Revised: November 2, 2014

Accepted: December 24, 2014

Published: January 29, 2015

REFERENCES

- Abraira, V.E., and Ginty, D.D. (2013). The sensory neurons of touch. *Neuron* 79, 618–639.
- Apps, R., and Garwicz, M. (2005). Anatomical and physiological foundations of cerebellar information processing. *Nat. Rev. Neurosci.* 6, 297–311.
- Arber, S. (2012). Motor circuits in action: specification, connectivity, and function. *Neuron* 74, 975–989.
- Armstrong, D.M. (1988). The supraspinal control of mammalian locomotion. *J. Physiol.* 405, 1–37.
- Basbaum, A.I., Bautista, D.M., Scherrer, G., and Julius, D. (2009). Cellular and molecular mechanisms of pain. *Cell* 139, 267–284.
- Bourane, S., Méchaly, I., Venteo, S., Garces, A., Fichard, A., Valmier, J., and Carroll, P. (2007). A SAGE-based screen for genes expressed in sub-populations of neurons in the mouse dorsal root ganglion. *BMC Neurosci.* 8, 97.
- Bretzner, F., and Drew, T. (2005). Motor cortical modulation of cutaneous reflex responses in the hindlimb of the intact cat. *J. Neurophysiol.* 94, 673–687.
- Bröhl, D., Strehle, M., Wende, H., Hori, K., Bormuth, I., Nave, K.A., Müller, T., and Birchmeier, C. (2008). A transcriptional network coordinately determines transmitter and peptidergic fate in the dorsal spinal cord. *Dev. Biol.* 322, 381–393.
- Brown, A.G. (1981). Organization in the spinal cord: the anatomy and physiology of identified neurones (Berlin, New York: Springer-Verlag).
- Buffelli, M., Burgess, R.W., Feng, G., Lobe, C.G., Lichtman, J.W., and Sanes, J.R. (2003). Genetic evidence that relative synaptic efficacy biases the outcome of synaptic competition. *Nature* 424, 430–434.
- Cheng, L., Arata, A., Mizuguchi, R., Qian, Y., Karunaratne, A., Gray, P.A., Arata, S., Shirasawa, S., Bouchard, M., Luo, P., et al. (2004). Tlx3 and Tlx1 are postmitotic selector genes determining glutamatergic over GABAergic cell fates. *Nat. Neurosci.* 7, 510–517.
- Chou, S.J., Babot, Z., Leingärtner, A., Studer, M., Nakagawa, Y., and O'Leary, D.D. (2013). Geniculocortical input drives genetic distinctions between primary and higher-order visual areas. *Science* 340, 1239–1242.
- Del Barrio, M.G., Bourane, S., Grossmann, K., Schüle, R., Britsch, S., O'Leary, D.D., and Goulding, M. (2013). A transcription factor code defines nine sensory interneuron subtypes in the mechanosensory area of the spinal cord. *PLoS ONE* 8, e77928.
- Delmas, P., Hao, J., and Rodat-Despoix, L. (2011). Molecular mechanisms of mechanotransduction in mammalian sensory neurons. *Nat. Rev. Neurosci.* 12, 139–153.
- Duan, B., Cheng, L., Bourane, S., Britz, O., Padilla, C., Garcia-Campmany, L., Krashes, M., Knowlton, W., Velasquez, T., Ren, X., et al. (2014). Identification of spinal circuits transmitting and gating mechanical pain. *Cell* 159, 1417–1432.
- Forsberg, H. (1979). Stumbling corrective reaction: a phase-dependent compensatory reaction during locomotion. *J. Neurophysiol.* 42, 936–953.

- Fyffe, R.E.W. (1992). Laminar organization of primary afferent terminals in the mammalian spinal cord. In *Sensory Neurons*, S.A. Scott, ed. (New York: Diversity, Development and Plasticity Oxford University Press), pp. 131–139.
- Goulding, M. (2009). Circuits controlling vertebrate locomotion: moving in a new direction. *Nat. Rev. Neurosci.* 10, 507–518.
- Grillner, S., and Jessell, T.M. (2009). Measured motion: searching for simplicity in spinal locomotor networks. *Curr. Opin. Neurobiol.* 19, 572–586.
- Grillner, S., Hongo, T., and Lund, S. (1971). Convergent effects on alpha motoneurons from the vestibulospinal tract and a pathway descending in the medial longitudinal fasciculus. *Exp. Brain Res.* 12, 457–479.
- Gross, M.K., Dottori, M., and Goulding, M. (2002). Lbx1 specifies somatosensory association interneurons in the dorsal spinal cord. *Neuron* 34, 535–549.
- Hantman, A.W., and Jessell, T.M. (2010). Clarke's column neurons as the focus of a corticospinal corollary circuit. *Nat. Neurosci.* 13, 1233–1239.
- Hu, J., Huang, T., Li, T., Guo, Z., and Cheng, L. (2012). c-Maf is required for the development of dorsal horn laminae III/IV neurons and mechanoreceptive DRG axon projections. *J. Neurosci.* 32, 5362–5373.
- Hultborn, H., Illert, M., and Santini, M. (1976). Convergence on interneurons mediating the reciprocal Ia inhibition of motoneurons. III. Effects from supraspinal pathways. *Acta Physiol. Scand.* 96, 368–391.
- Ito, M. (2012). *The cerebellum: brain for an implicit self* (Upper Saddle River, N.J.: FT Press).
- Jankowska, E. (1992). Interneuronal relay in spinal pathways from proprioceptors. *Prog. Neurobiol.* 38, 335–378.
- Kiehn, O. (2011). Development and functional organization of spinal locomotor circuits. *Curr. Opin. Neurobiol.* 21, 100–109.
- Koch, S.C., Tochiki, K.K., Hirschberg, S., and Fitzgerald, M. (2012). C-fiber activity-dependent maturation of glycinergic inhibition in the spinal dorsal horn of the postnatal rat. *Proc. Natl. Acad. Sci. USA* 109, 12201–12206.
- Lallemend, F., and Ernfor, P. (2012). Molecular interactions underlying the specification of sensory neurons. *Trends Neurosci.* 35, 373–381.
- Lawson, S.N., Crepps, B., and Perl, E.R. (2002). Calcitonin gene-related peptide immunoreactivity and afferent receptive properties of dorsal root ganglion neurons in guinea-pigs. *J. Physiol.* 540, 989–1002.
- Lechner, S.G., and Lewin, G.R. (2013). Hair sensation. *Physiology (Bethesda)* 28, 142–150.
- Li, L., Rutlin, M., Abaira, V.E., Cassidy, C., Kus, L., Gong, S., Jankowski, M.P., Luo, W., Heintz, N., Koerber, H.R., et al. (2011). The functional organization of cutaneous low-threshold mechanosensory neurons. *Cell* 147, 1615–1627.
- Li, Y., Stam, F.J., Aimone, J.B., Goulding, M., Callaway, E.M., and Gage, F.H. (2013). Molecular layer perforant path-associated cells contribute to feed-forward inhibition in the adult dentate gyrus. *Proc. Natl. Acad. Sci. USA* 110, 9106–9111.
- Light, A.R., and Perl, E.R. (1979a). Reexamination of the dorsal root projection to the spinal dorsal horn including observations on the differential termination of coarse and fine fibers. *J. Comp. Neurol.* 186, 117–131.
- Light, A.R., and Perl, E.R. (1979b). Spinal termination of functionally identified primary afferent neurons with slowly conducting myelinated fibers. *J. Comp. Neurol.* 186, 133–150.
- Liu, Q., Vrontou, S., Rice, F.L., Zylka, M.J., Dong, X., and Anderson, D.J. (2007). Molecular genetic visualization of a rare subset of unmyelinated sensory neurons that may detect gentle touch. *Nat. Neurosci.* 10, 946–948.
- Lundberg, A., and Voorhoeve, P. (1962). Effects from the pyramidal tract on spinal reflex arcs. *Acta Physiol. Scand.* 56, 201–219.
- Madisen, L., Zwingman, T.A., Sunkin, S.M., Oh, S.W., Zariwala, H.A., Gu, H., Ng, L.L., Palmiter, R.D., Hawrylycz, M.J., Jones, A.R., et al. (2010). A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* 13, 133–140.
- McGlone, F., and Reilly, D. (2010). The cutaneous sensory system. *Neurosci. Biobehav. Rev.* 34, 148–159.
- Miles, G.B., Hartley, R., Todd, A.J., and Brownstone, R.M. (2007). Spinal cholinergic interneurons regulate the excitability of motoneurons during locomotion. *Proc. Natl. Acad. Sci. USA* 104, 2448–2453.
- Müller, T., Brohmann, H., Pierani, A., Heppenstall, P.A., Lewin, G.R., Jessell, T.M., and Birchmeier, C. (2002). The homeodomain factor *lhx1* distinguishes two major programs of neuronal differentiation in the dorsal spinal cord. *Neuron* 34, 551–562.
- Patel, M., Magnusson, M., Kristinsdottir, E., and Fransson, P.A. (2009). The contribution of mechanoreceptive sensation on stability and adaptation in the young and elderly. *Eur. J. Appl. Physiol.* 105, 167–173.
- Perry, S.D., McLroy, W.E., and Maki, B.E. (2000). The role of plantar cutaneous mechanoreceptors in the control of compensatory stepping reactions evoked by unpredictable, multi-directional perturbation. *Brain Res.* 877, 401–406.
- Prätorius, B., Kimmekamp, S., and Milani, T.L. (2003). The sensitivity of the sole of the foot in patients with Morbus Parkinson. *Neurosci. Lett.* 346, 173–176.
- Quevedo, J., Stecina, K., Gosgnach, S., and McCrea, D.A. (2005). Stumbling corrective reaction during fictive locomotion in the cat. *J. Neurophysiol.* 94, 2045–2052.
- Rossignol, S., Dubuc, R., and Gossard, J.P. (2006). Dynamic sensorimotor interactions in locomotion. *Physiol. Rev.* 86, 89–154.
- Sakurai, K., Akiyama, M., Cai, B., Scott, A., Han, B.X., Takatoh, J., Sigrist, M., Arber, S., and Wang, F. (2013). The organization of submodality-specific touch afferent inputs in the vibrissa column. *Cell Rep.* 5, 87–98.
- Schouenborg, J. (2008). Action-based sensory encoding in spinal sensorimotor circuits. *Brain Res. Brain Res. Rev.* 57, 111–117.
- Seidler, B., Schmidt, A., Mayr, U., Nakhai, H., Schmid, R.M., Schneider, G., and Saur, D. (2008). A Cre-loxP-based mouse model for conditional somatic gene expression and knockdown in vivo by using avian retroviral vectors. *Proc. Natl. Acad. Sci. USA* 105, 10137–10142.
- Semba, K., Masarachia, P., Malamed, S., Jacquin, M., Harris, S., and Egger, M.D. (1984). Ultrastructure of pacinian corpuscle primary afferent terminals in the cat spinal cord. *Brain Res.* 302, 135–150.
- Shortland, P., and Woolf, C.J. (1993). Morphology and somatotopy of the central arborizations of rapidly adapting glabrous skin afferents in the rat lumbar spinal cord. *J. Comp. Neurol.* 329, 491–511.
- Stål, F., Fransson, P.A., Magnusson, M., and Karlberg, M. (2003). Effects of hypothermic anesthesia of the feet on vibration-induced body sway and adaptation. *J. Vestib. Res.* 13, 39–52.
- Stam, F.J., Hendricks, T.J., Zhang, J., Geiman, E.J., Francius, C., Labosky, P.A., Clotman, F., and Goulding, M. (2012). Renshaw cell interneuron specialization is controlled by a temporally restricted transcription factor program. *Development* 139, 179–190.
- Stepien, A.E., Tripodi, M., and Arber, S. (2010). Monosynaptic rabies virus reveals premotor network organization and synaptic specificity of cholinergic partition cells. *Neuron* 68, 456–472.
- Todd, A.J. (2010). Neuronal circuitry for pain processing in the dorsal horn. *Nat. Rev. Neurosci.* 11, 823–836.
- Torsney, C., and MacDermott, A.B. (2006). Disinhibition opens the gate to pathological pain signaling in superficial neurokinin 1 receptor-expressing neurons in rat spinal cord. *J. Neurosci.* 26, 1833–1843.
- Uesaka, T., Nagashimada, M., Yonemura, S., and Enomoto, H. (2008). Diminished Ret expression compromises neuronal survival in the colon and causes intestinal aganglionosis in mice. *J. Clin. Invest.* 118, 1890–1898.
- Vrontou, S., Wong, A.M., Rau, K.K., Koerber, H.R., and Anderson, D.J. (2013). Genetic identification of C fibres that detect massage-like stroking of hairy skin in vivo. *Nature* 493, 669–673.
- Wang, X., Zhang, J., Eberhart, D., Urban, R., Meda, K., Solorzano, C., Yamana, H., Rice, D., and Basbaum, A.I. (2013). Excitatory superficial dorsal horn interneurons are functionally heterogeneous and required for the full behavioral expression of pain and itch. *Neuron* 78, 312–324.
- Wickersham, I.R., Lyon, D.C., Barnard, R.J., Mori, T., Finke, S., Conzelmann, K.K., Young, J.A., and Callaway, E.M. (2007). Monosynaptic restriction of

transsynaptic tracing from single, genetically targeted neurons. *Neuron* 53, 639–647.

Woodbury, C.J., Ritter, A.M., and Koerber, H.R. (2001). Central anatomy of individual rapidly adapting low-threshold mechanoreceptors innervating the “hairy” skin of newborn mice: early maturation of hair follicle afferents. *J. Comp. Neurol.* 436, 304–323.

Woolf, C.J. (1987). Central terminations of cutaneous mechanoreceptive afferents in the rat lumbar spinal cord. *J. Comp. Neurol.* 261, 105–119.

Xu, Y., Lopes, C., Wende, H., Guo, Z., Cheng, L., Birchmeier, C., and Ma, Q. (2013). Ontogeny of excitatory spinal neurons processing distinct somatic sensory modalities. *J. Neurosci.* 33, 14738–14748.

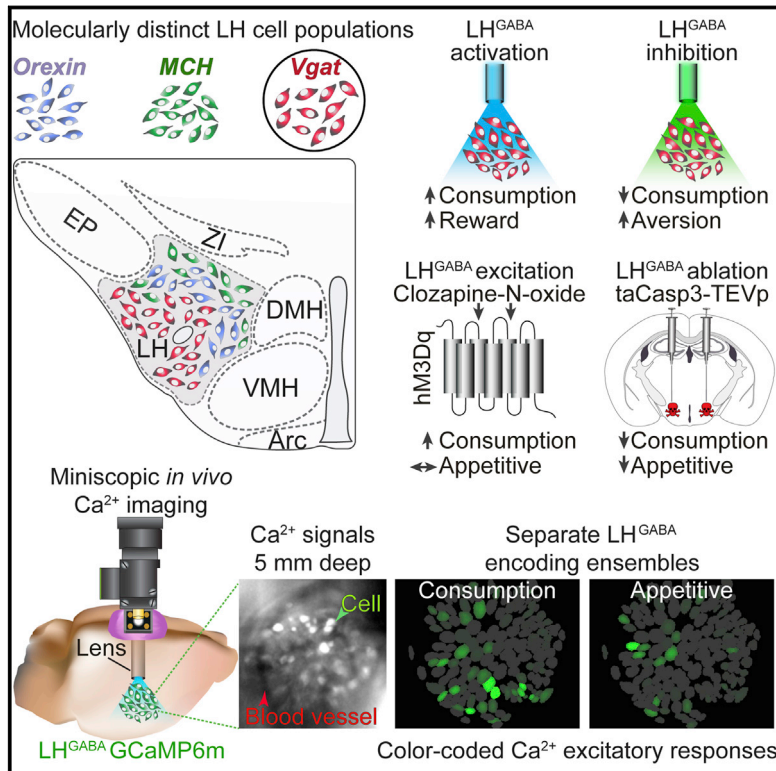
Zagoraoui, L., Akay, T., Martin, J.F., Brownstone, R.M., Jessell, T.M., and Miles, G.B. (2009). A cluster of cholinergic premotor interneurons modulates mouse locomotor activity. *Neuron* 64, 645–662.

Zhang, J., Lanuza, G.M., Britz, O., Wang, Z., Siembab, V.C., Zhang, Y., Velasquez, T., Alvarez, F.J., Frank, E., and Goulding, M. (2014). V1 and v2b interneurons secure the alternating flexor-extensor motor activity mice require for limbed locomotion. *Neuron* 82, 138–150.

Zia, S., Cody, F.W., and O’Boyle, D.J. (2003). Discrimination of bilateral differences in the loci of tactile stimulation is impaired in subjects with Parkinson’s disease. *Clin. Anat.* 16, 241–247.

Visualizing Hypothalamic Network Dynamics for Appetitive and Consummatory Behaviors

Graphical Abstract



Authors

Joshua H. Jennings, Randall L. Ung, ..., Stephani Otte, Garret D. Stuber

Correspondence

gstuber@med.unc.edu

In Brief

Deep brain calcium imaging in freely behaving mice reveals that appetitive and consummatory behaviors are encoded in distinct neurons in the hypothalamus, suggesting the existence of separate networks regulating motivation to eat and food consumption.

Highlights

- Activation of LH GABAergic neurons produces appetitive and consummatory behaviors
- Ablation of LH GABAergic cells attenuates weight gain, consumption, and motivation
- LH GABAergic neuronal subsets are molecularly distinct from MCH and Orx cells
- *In vivo* Ca²⁺ imaging reveals separate appetitive and consummatory encoding networks



Visualizing Hypothalamic Network Dynamics for Appetitive and Consummatory Behaviors

Joshua H. Jennings,^{1,2,3,6} Randall L. Ung,^{1,2,3,6} Shanna L. Resendez,^{1,3} Alice M. Stamatakis,^{1,2,3} Johnathon G. Taylor,¹ Jonathan Huang,¹ Katie Veleta,² Pranish A. Kantak,¹ Megumi Aita,¹ Kelson Shilling-Scriver,¹ Charu Ramakrishnan,⁴ Karl Deisseroth,⁴ Stephani Otte,⁵ and Garret D. Stuber^{1,2,*}

¹Departments of Psychiatry and Cell Biology and Physiology, University of North Carolina, Chapel Hill, NC 27599, USA

²Curriculum in Neurobiology, University of North Carolina, Chapel Hill, NC 27599, USA

³Neuroscience Center, University of North Carolina, Chapel Hill, NC 27599, USA

⁴Department of Bioengineering, Stanford University, Stanford, CA 94305, USA

⁵Inscopix Inc., Palo Alto, CA 94303, USA

⁶Co-first author

*Correspondence: gstuber@med.unc.edu

<http://dx.doi.org/10.1016/j.cell.2014.12.026>

SUMMARY

Optimally orchestrating complex behavioral states, such as the pursuit and consumption of food, is critical for an organism's survival. The lateral hypothalamus (LH) is a neuroanatomical region essential for appetitive and consummatory behaviors, but whether individual neurons within the LH differentially contribute to these interconnected processes is unknown. Here, we show that selective optogenetic stimulation of a molecularly defined subset of LH GABAergic (*Vgat*-expressing) neurons enhances both appetitive and consummatory behaviors, whereas genetic ablation of these neurons reduced these phenotypes. Furthermore, this targeted LH subpopulation is distinct from cells containing the feeding-related neuropeptides, melanin-concentrating hormone (MCH), and orexin (Orx). Employing in vivo calcium imaging in freely behaving mice to record activity dynamics from hundreds of cells, we identified individual LH GABAergic neurons that preferentially encode aspects of either appetitive or consummatory behaviors, but rarely both. These tightly regulated, yet highly intertwined, behavioral processes are thus dissociable at the cellular level.

INTRODUCTION

Complex, but evolutionary well-conserved, behavioral states, such as those related to feeding, require both precurrent and consummatory responses (Sherrington, 1906). These oftentimes sequential reactions, which are historically conceptualized as appetitive and consummatory behaviors (Ball and Balthazart, 2008; Craig, 1917; Lorenz, 1950; Tinbergen, 1951), represent highly intertwined processes that are not fully distinguished at

the neural level. The lateral hypothalamus (LH), a critical modulator of both appetitive and consummatory processes, is a heterogeneous brain area containing numerous genetically distinct cell populations that utilize various signaling modalities (Berthoud and Münzberg, 2011). Gene expression patterns within the LH suggest that individual neurons likely release either inhibitory or excitatory neurotransmitters, γ -aminobutyric acid (GABA), and glutamate, as well as a host of neuropeptides (Hökfelt et al., 2000; Lein et al., 2007), implying that identifiable subdivisions within the global LH neuronal network can be genetically targeted. Electrical stimulation of the LH, which nonspecifically activates various cell types and processes, evokes voracious consummatory feeding, as well as appetitive reward-related behaviors (Hoebel and Teitelbaum, 1962; Margules and Olds, 1962; Olds and Milner, 1954; Wise, 1971), whereas ablation of the region results in emaciation and aphagia (Anand and Brobeck, 1951; Hoebel, 1965). Moreover, the activity of LH neurons changes in response to food and associated stimuli (Ono et al., 1981, 1986). These electrically evoked behaviors and feeding-specific activity patterns were discovered in numerous species of the Kingdom Animalia, from lizards (Molina-Borja and Gómez-Soutullo, 1989) to humans (Quaade et al., 1974). These past findings emphasize the importance and evolutionary conservation of the LH for regulating these survival-oriented behaviors. However, given the heterogeneous cellular composition of the LH (Adamantidis and de Lecea, 2009; Kamani et al., 2013; Knight et al., 2012), and the fact that multiple fibers of passage traverse this region (Hahn and Swanson, 2012), classical electrical stimulation, lesion, or electrophysiological recording studies are unable to determine whether genetically defined cell types, such as LH GABAergic neurons, encode, and regulate precise aspects of appetitive food-seeking and/or consummatory behaviors.

Neural circuit tracing and manipulation experiments revealed that optogenetic modulation of LH glutamatergic neurons influences feeding and motivated behavioral responding (Jennings et al., 2013a). In the current study, we examined if molecularly defined LH neurons that express the gene for the vesicular GABA transporter (*Vgat*), and thus synthesize and release

GABA, selectively promote and encode appetitive and consummatory behaviors.

RESULTS

Optogenetic Stimulation and Inhibition of LH GABAergic Neurons Bidirectionally Modulate Feeding and Reward-Seeking Behavior

Given that electrical stimulation of the LH produces multiple behavioral responses, including feeding and positive reinforcement (Hoebel and Teitelbaum, 1962; Margules and Olds, 1962), and that the highest concentration of GABA in the hypothalamus is located within the anterior portions of the LH (Kimura and Kuriyama, 1975), we first explored the functional role of LH GABAergic neurons in modulating feeding and reward-related behaviors by using optogenetic techniques. Applying established viral procedures (Jennings et al., 2013b), we first expressed channelrhodopsin-2 conjugated to enhanced yellow fluorescent protein (AAV5-EF1 α -DIO-hChR2(H134R)-eYFP) selectively in GABAergic neurons located within the anterior and dorsolateral subdivisions of the LH (Figures 1A and 1B) in *Vgat-IRES-Cre* mice (Vong et al., 2011) and implanted optical fibers directly above the LH for somata photostimulation (Figure 1C; Figure S1A available online). Approximately 4 weeks after surgery, we tested whether direct photostimulation of LH GABAergic neurons influenced feeding and reward-related behavioral phenotypes in ad-libitum-fed mice. Photoactivation of these neurons at 20 Hz significantly increased the time spent in a designated food area (Figures 1D and 1E), food consumption (Figure 1F), time spent in a location paired with photostimulation (Figure 1G), and optical self-stimulation behavior (Figure 1H). Next, we tested whether photoinhibition of LH GABAergic neurons disrupted feeding and reward-related behaviors in food-restricted mice. Utilizing similar procedures as described above, we targeted a modified variant (Mattis et al., 2012) of the inhibitory opsin, archaerhodopsin-3 (AAV5-EF1 α -DIO-eArch3.0-eYFP; Chow et al., 2010), to LH GABAergic neurons in *Vgat-IRES-Cre* mice (Figures 1I–1K; Figure S1B). Photoinhibition of LH GABAergic neurons led to significant reductions in time spent in a designated food area (Figures 1L and 1M), food consumption (Figure 1N), and time spent in a location paired with photoinhibition (Figure 1O). These data indicate that selective optogenetic modulation of neurochemically distinct LH GABAergic neurons (Figures S1C–S1Q) influences both feeding and reward-related phenotypes, supporting the idea that both appetitive and consummatory behavioral processes are represented in the LH by GABAergic neurons.

Chemogenetic Activation of LH GABAergic Neurons Enhances Consummatory Behaviors

To expand upon the acute behavioral effects observed from optogenetic manipulations (Figure 1), we investigated if sustained activation of LH GABAergic neurons, over a longer timescale, influenced consumption and work effort to earn a caloric reward. Thus, we virally targeted the G $_q$ -coupled excitatory designer receptor exclusively activated by designer drugs (DREADD), hM3Dq, to LH GABAergic neurons by injecting the Cre-inducible viral construct, AAV8-hSyn-DIO-hM3D(Gq)-mCherry (Krashes

et al., 2011), into similar target zones within the LH of *Vgat-IRES-Cre* mice (LH^{GABA::hM3Dq}; Figures 2A and 2B; Figures S2A–S2G). The inert molecule, clozapine-N-oxide (CNO), selectively binds to hM3Dq and activates neurons through G $_q$ signaling pathways (Alexander et al., 2009). Therefore, to verify CNO-mediated activation in LH^{GABA::hM3Dq} neurons, we performed whole-cell recordings in brain slices and found that the spontaneous firing rate of a subset of LH^{GABA::hM3Dq} neurons significantly increased upon CNO (5 μ M) bath application (Figures 2C and 2D). Additionally, we examined whether in vivo stimulation of LH GABAergic neurons, via systemic CNO administration, enhanced the expression of Fos, a marker for neuronal activity, within the LH. Intraperitoneal (i.p.) injections of CNO (1 mg/kg) in LH^{GABA::hM3Dq} mice significantly increased Fos expression in the LH compared to LH^{GABA::Control} mice injected with CNO (Figures 2E–2I). To determine whether DREADD-mediated activation of LH GABAergic neurons affected consummatory behavior, mice were trained in a free-access caloric consumption task (1-hr duration) that permitted the quantification of lick responses at the delivery spout for a palatable caloric liquid reward. Chemogenetic activation of LH GABAergic neurons in LH^{GABA::hM3Dq} mice via CNO (1 mg/kg, i.p.) injection led to a significant increase in the number of lick (consummatory) responses when compared to controls (Figures 2J and 2K). In addition, CNO administration 45 min prior to a 1 hr free-access feeding task (Figure 2L) enhanced food intake in ad-libitum-fed LH^{GABA::hM3Dq} mice (Figure 2M; see the Extended Experimental Procedures).

To examine whether chemogenetic activation of LH GABAergic neurons alters work effort exerted to earn a caloric reward, we initially trained mice on a fixed ratio of one (FR1) schedule, where each active nose poke resulted in the delivery of a palatable and calorie-dense liquid. Following stable behavioral responding, on subsequent sessions, LH^{GABA::hM3Dq} and LH^{GABA::Control} mice were either administered saline or CNO (counterbalanced) 45 min prior to progressive ratio 3 (PR3) test sessions, which is an established behavioral assay for measuring an animal's motivation to obtain caloric rewards (Krashes et al., 2011). Intriguingly, activation of these cells significantly increased lick responses (metrics of consumption; Figures 2N and 2O), but did not alter the number of active nose pokes nor the break point (metrics of motivation; Figures 2P and 2Q). Anecdotally, we observed that LH^{GABA::hM3Dq} mice treated with CNO spent majority of the time licking at the reward receptacle (despite reward delivery being contingent upon nose poke responding), suggesting that bulk chemogenetic activation of these neurons may override optimal behavioral performance by biasing behavior toward aspects of consumption. Taken together, these data indicate that chemogenetic activation of LH GABAergic neurons enhances consummatory behaviors.

Vgat-Targeted LH GABAergic Neurons Are Molecularly Distinct from MCH and Orx Cells

The LH exclusively houses two separate molecularly defined cell types, melanin-concentrating hormone (MCH) and orexin (Orx) neurons (Elias et al., 1998), which are known to have diverse roles in regulating food intake, energy balance, reward, and

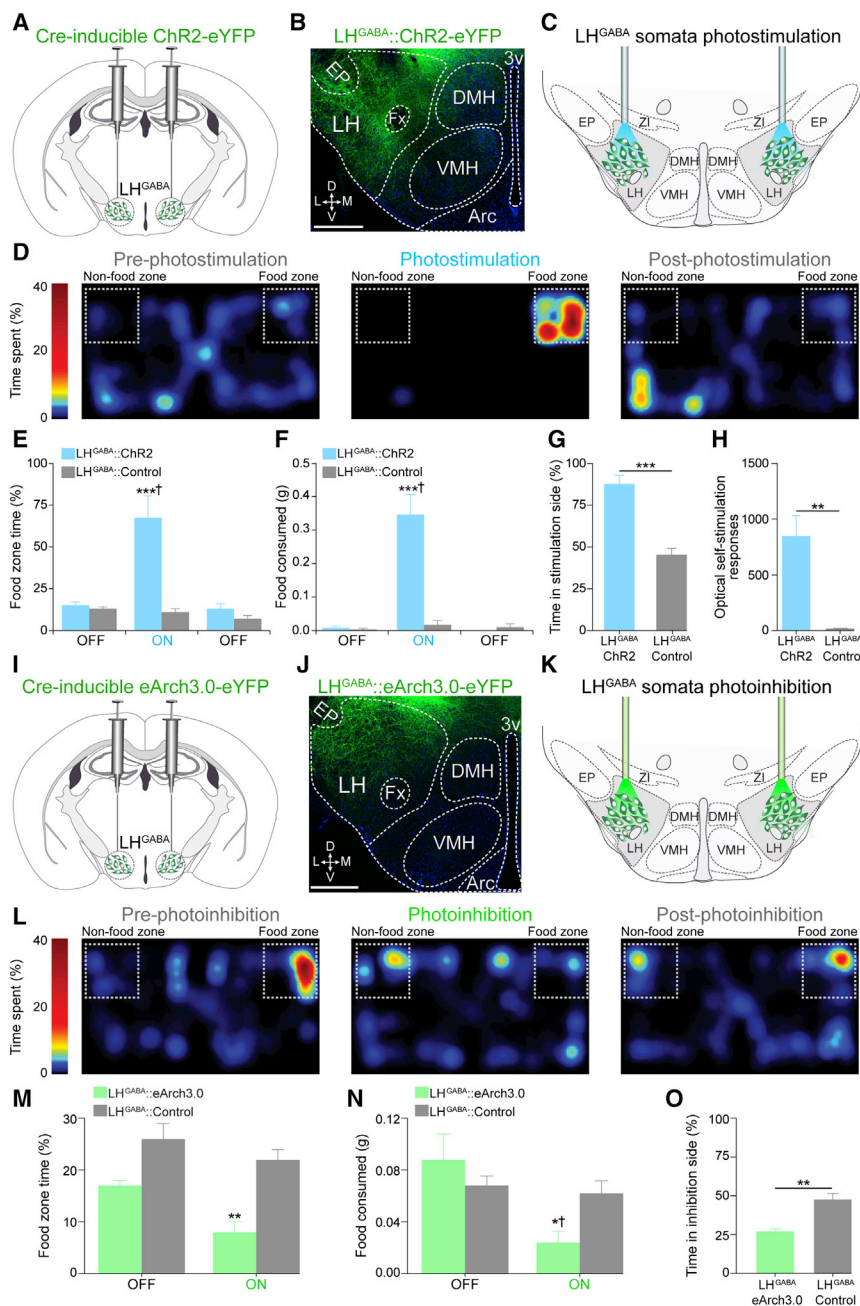


Figure 1. Optogenetic Modulation of LH GABAergic Neurons Bidirectionally Modulates Feeding and Reward-Related Behaviors

(A) Scheme for viral targeting of AAV5-EF1 α -DIO-ChR2-eYFP to the LH of *Vgat-IRES-Cre* animals. (B) 20x confocal image depicting ChR2-eYFP expression in LH GABAergic neurons. EP, entopeduncular nucleus; LH, lateral hypothalamus; Fx, fornix; DMH, dorsomedial hypothalamic nucleus; VMH, ventromedial hypothalamic nucleus; Arc, arcuate nucleus; 3v, third ventricle; D, dorsal; L, lateral; M, medial; V, ventral. Scale bar, 200 μ m. (C) Diagram for photostimulation of LH GABAergic neurons.

(D) Color map encoding spatial location from an ad-lib-fed LH^{GABA::}ChR2 mouse during the free-access feeding paradigm.

(E) Photostimulation of LH GABAergic neurons significantly increased time spent in the food zone compared to controls and time epochs without photostimulation ($n = 5$ mice per group; $F_{2,27} = 86.24$; $p < 0.0001$).

(F) 20 Hz photostimulation significantly increased food intake in LH^{GABA::}ChR2 mice compared to controls and time epochs without photostimulation ($n = 5$ mice per group; $F_{2,27} = 17.05$; $p < 0.0001$).

(G) LH^{GABA::}ChR2 mice spent significantly more time in the chamber paired with photostimulation compared to controls ($n = 5$ mice per group; $t_8 = 6.796$; $p < 0.0001$).

(H) LH^{GABA::}ChR2 mice nose poked significantly more for 20 Hz photostimulation compared to controls ($n = 4$ mice per group; $t_5 = 5.744$; $p = 0.0012$).

(I) Viral targeting of AAV5-EF1 α -DIO-eArch3.0-eYFP into the LH of *Vgat-IRES-Cre* mice.

(J) 20x confocal image showing eArch3.0-eYFP expression in the LH of a *Vgat-IRES-Cre* mouse. Scale bar, 200 μ m.

(K) Illustration for somata LH GABAergic photoinhibition.

(L) Color map encoding spatial location of an example food-deprived LH^{GABA::}eArch3.0 animal during the free-access feeding paradigm.

(M) Upon photoinhibition exposure, LH^{GABA::}eArch3.0 animals spent significantly less time in the food zone compared to controls ($n = 5$ mice per group; $F_{1,16} = 9.39$; $p = 0.007$).

(N) Photoinhibition of LH GABAergic neurons significantly suppressed food intake in food-restricted mice when compared to controls and time epochs without photoinhibition ($n = 5$ mice per group; $F_{1,16} = 5.43$; $p = 0.033$).

(O) LH^{GABA::}eArch3.0 mice spent significantly less time in the photoinhibition-paired chamber compared to controls ($n = 5$ mice per group; $t_8 = 4.512$; $p = 0.002$). All values are mean \pm SEM. Student's t test or two-way ANOVA; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. See also Figure S1.

sleep-wakefulness (Jego et al., 2013; Karnani et al., 2011; Sakurai et al., 1998; Whiddon and Palmiter, 2013). Because MCH and Orx are thought to promote aspects of feeding and reward-related behaviors, we next assessed whether *Vgat*-targeted neurons within the LH coexpress either of these orexigenic neuropeptides. To achieve this, we immunolabeled both peptide-producing neuronal populations in *Vgat-IRES-Cre* mice that endogenously express eYFP in *Vgat*⁺ cells (*Vgat*-eYFP)

and determined whether Orx and/or MCH neurons colocalize with *Vgat*⁺ cells in the LH. Strikingly, immunostaining for MCH and Orx in *Vgat*-eYFP (*Vgat-IRES-Cre* mice crossed to a Ai3 reporter line; Madisen et al., 2010) brain slices revealed that *Vgat*⁺ LH neurons do not coexpress either of these neuropeptides, signifying that these *Vgat*-targeted neurons in the LH represent a neurochemically distinct GABAergic subpopulation that is separate from MCH and Orx cells (Figures 3A–3F; Figure S3).

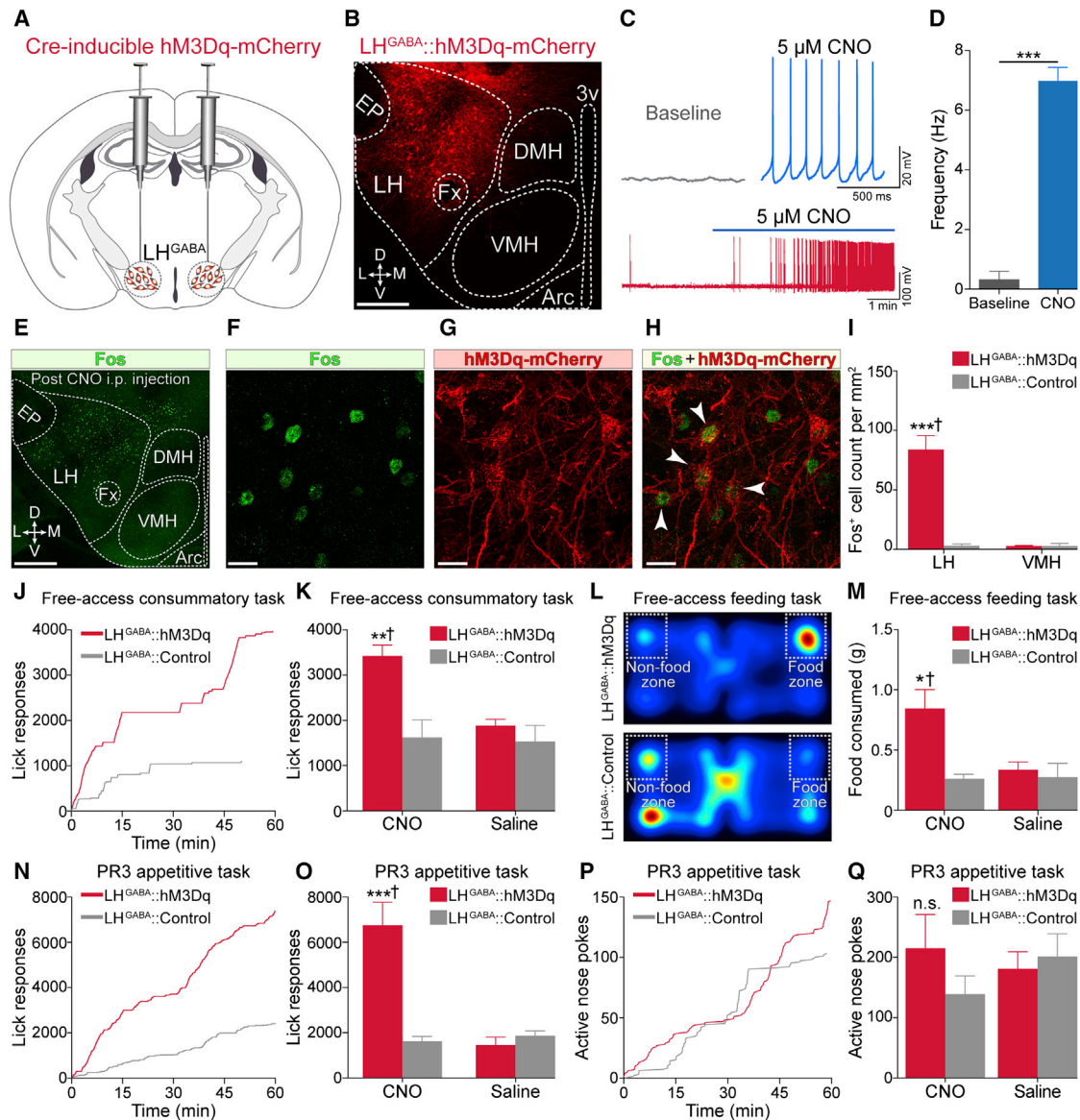


Figure 2. Bulk Chemogenetic Activation of LH GABAergic Neurons Enhances Consummatory Behaviors

(A) Viral targeting of AAV8-hSyn-DIO-hM3D(Gq)-mCherry into the LH of *Vgat-IRES-Cre* mice.

(B) 20x confocal image showing hM3Dq-mCherry expression in the LH of a *Vgat-IRES-Cre* animal. Scale bar, 200 μ m.

(C) Example current-clamp traces from a LH^{GABA::hM3Dq} brain slice before (baseline) and after 5 μ M CNO demonstrating DREADD-mediated action potentials (bottom).

(D) CNO significantly increases spontaneous firing of LH^{GABA::hM3Dq} neurons ($n = 3$ cells; $n = 3$ mice; $t_4 = 12.370$; $p = 0.0002$).

(E) 20x confocal image from an example LH^{GABA::hM3Dq} animal, where Fos-induction was assessed 2 hr after CNO (1 mg/kg; i.p.) injection. Scale bar, 500 μ m.

(F–H) 63x confocal images displaying Fos immunoreactivity (F) and hM3Dq-mCherry expression (G) in a LH^{GABA::hM3Dq} animal injected with CNO. (H) Merged image of (F) and (G) showing colocalization of Fos and hM3Dq-mCherry expression as indicated by white arrows. Scale bars, 20 μ m.

(I) CNO administration in LH^{GABA::hM3Dq} mice significantly increases Fos expression in the LH compared to controls and neighboring regions ($n = 3$ LH sections; $n = 3$ mice per group; $F_{1,8} = 46.00$; $p < 0.0001$).

(J) Cumulative lick responses from individual LH^{GABA::hM3Dq} and LH^{GABA::Control} mice during a single 1 hr free-access caloric consumption task.

(K) Systemic CNO application significantly increases lick responses in LH^{GABA::hM3Dq} mice compared to LH^{GABA::Control} mice and saline injections during a free-access caloric consumption task ($n = 6$ mice per group; $F_{1,20} = 8.12$; $p = 0.01$).

(L) Example color maps from LH^{GABA::hM3Dq} (top) and LH^{GABA::Control} mice (bottom) during the free-access feeding task.

(M) Systemic application of CNO in LH^{GABA::hM3Dq} mice significantly increased food consumption during the free-access feeding task when compared to controls and saline injections ($n = 6$ mice per group; $F_{1,20} = 6.37$; $p = 0.02$).

(N) Cumulative lick responses from example LH^{GABA::hM3Dq} and LH^{GABA::Control} animals during a single 1 hr progressive ratio 3 (PR3) session.

(legend continued on next page)

Although these neuropeptide-producing neurons have been implicated as crucial modulators of feeding, selective activation of *Vgat*-targeted neurons within the LH does not directly stimulate MCH and Orx cells to produce feeding and reward-related behaviors (Figures S1F–S1Q and S2B–S2G).

Genetic Ablation of LH GABAergic Neurons Reduces Consummatory Behaviors and Motivation to Obtain Caloric Rewards

Next, we assessed the necessity of LH GABAergic neurons for regulating these feeding-related processes with cell-type-specific ablation methods. To selectively ablate LH GABAergic neurons, we injected a Cre-inducible viral construct coding for taCasp3-2A-TEVp (AAV2-FLEX-taCasp3-TEVp) into the LH of *Vgat-IRES-Cre* mice (LH^{GABA::taCasp3}; Figure 3G) (Yang et al., 2013). In situ hybridization for glutamic acid decarboxylase (*GAD67*), a marker for GABAergic neurons (Erlander et al., 1991), revealed a significant reduction in the number of LH GABAergic neurons following bilateral LH injections of AAV-FLEX-taCasp3-TEVp, and no change in the number of GABAergic neurons within neighboring regions, including the entopeduncular nucleus (EP) and ventromedial hypothalamus (VMH; Figures 3H–3J). Previous studies demonstrated that chemical or electrolytic lesions of the LH can result in significant aphagia and weight loss (Harrell et al., 1975; Schallert and Whishaw, 1978). Therefore, we monitored the daily body weight of LH^{GABA::taCasp3} and LH^{GABA::Control} mice, while both groups were maintained on a calorie-dense diet for 60 days. Ablation of LH GABAergic neurons significantly blunted weight gain in LH^{GABA::taCasp3} mice compared to controls (Figure 3K). Additionally, LH^{GABA::taCasp3} mice displayed a significant reduction in daily food intake measured at 1 month postvirus injection (Figure 3L).

Next, we tested the effects of LH GABAergic neuron ablation on acute feeding and motivation to obtain caloric reward. Food-restricted LH^{GABA::taCasp3} mice showed reduced food intake in the acute free-access feeding assay (Figures 3M and 3N) and diminished lick responding in the free-access caloric consumption task compared to controls (Figure 3O). When tested on the PR3 task, LH^{GABA::taCasp3} mice exhibited reduced lick responding compared to controls (Figures 3P and 3Q). Congruently, LH^{GABA::taCasp3} mice showed a significant reduction in metrics of appetitive responding (active nose pokes and break points) compared to controls (Figures 3R–3T). Furthermore, genetic ablation of LH GABAergic neurons did not significantly alter the number of MCH or Orx cells in the LH (Figures S4A–S4F), corroborating our findings that *Vgat*-targeted LH neurons are a separate population from MCH and Orx cells (Figures 3A–3F; Figure S3). Lastly, LH^{GABA::taCasp3} mice did not display locomotor deficits or anxiety-related phenotypes when tested in an open-field test (Figures S4G and S4H). Together, these data highlight a causal and specific role for LH

GABAergic neurons in regulating appetitive responding, consumption, and energy balance.

In Vivo Ca²⁺ Imaging in Large Populations of LH GABAergic Neurons

The optogenetic, chemogenetic, and cell ablation experiments outlined above demonstrate that bulk modulation of LH GABAergic neurons causally regulates appetitive and consummatory behavioral output. Though these approaches provide important causal information for linking cell-type-specific function to behavior, bulk modulation of even genetically defined neurons fails to account for the high degree of functional diversity within a targeted cell population, and they do not accurately reflect endogenous cellular activity dynamics that underlie complex behaviors. Therefore, it is unclear whether appetitive and consummatory processes are orchestrated by functionally discrete LH GABAergic subpopulations or whether individual neurons are participant in aspects of both appetitive and consummatory behaviors. To examine this in detail, we applied in vivo microendoscopic imaging strategies (Barretto et al., 2011; Flusberg et al., 2008; Ghosh et al., 2011; Jennings and Stuber, 2014) to resolve somatic Ca²⁺ activity dynamics from hundreds of LH GABAergic neurons ($n = 743$ cells) in freely behaving mice ($n = 6$ mice; Movie S1). First, we expressed the green fluorescent sensor of neuronal activity, GCaMP6m, in LH GABAergic neurons by injecting a Cre-inducible AAV viral construct (AAVDJ-EF1 α -DIO-GCaMP6m) into the LH of *Vgat-IRES-Cre* mice (LH^{GABA::GCaMP6m}; Figures 4A–4C; Figures S5A–S5L). To circumvent the optical aberrations and the turbidity of tissue that typically preclude imaging in deep brain structures like the LH (~5 mm deep), we implanted 8-mm-long microendoscopes (0.5 mm diameter; consisting of a relay lens fused to doublet gradient refractive index microlenses) directly above the LH for optical detection of GCaMP6m fluorescence emission (Figures 4D–4F; Figure S5M). Next, we interfaced the implanted microendoscope with a detachable miniaturized fluorescence microscope (Ghosh et al., 2011; Ziv et al., 2013) to visualize Ca²⁺ signals from large populations of LH GABAergic neurons in freely moving LH^{GABA::GCaMP6m} mice (Figure 4G). Furthermore, applying this technique in conjunction with established computational algorithms (Mukamel et al., 2009), we were able to identify and track the Ca²⁺ activity from individual LH GABAergic neurons within single recording sessions and across multiple days and behavioral tasks (Figures 4H and 4I; see the Extended Experimental Procedures).

Food-Location Coding Profiles of LH GABAergic Neurons

Because the activity of unclassified LH neurons can be modulated in response to nutrient-related information (Ono et al., 1981, 1986), we first determined whether LH GABAergic ensembles are engaged when food is readily available. Thus, we

(O) Systemic CNO application significantly increased lick responses in LH^{GABA::hM3Dq} mice during the PR3 paradigm when compared to controls and saline injections ($n = 6$ mice per group; $F_{1,20} = 24.37$; $p < 0.0001$).

(P) Nose poke responses from example LH^{GABA::hM3Dq} and LH^{GABA::Control} mice during a single PR3 session.

(Q) CNO administration did not significantly affect nose poke responses during the PR3 session ($n = 6$ mice per group; $F_{1,20} = 1.47$; $p = 0.24$).

All values are mean \pm SEM. Student's *t* test or two-way ANOVA. See also Figure S2.

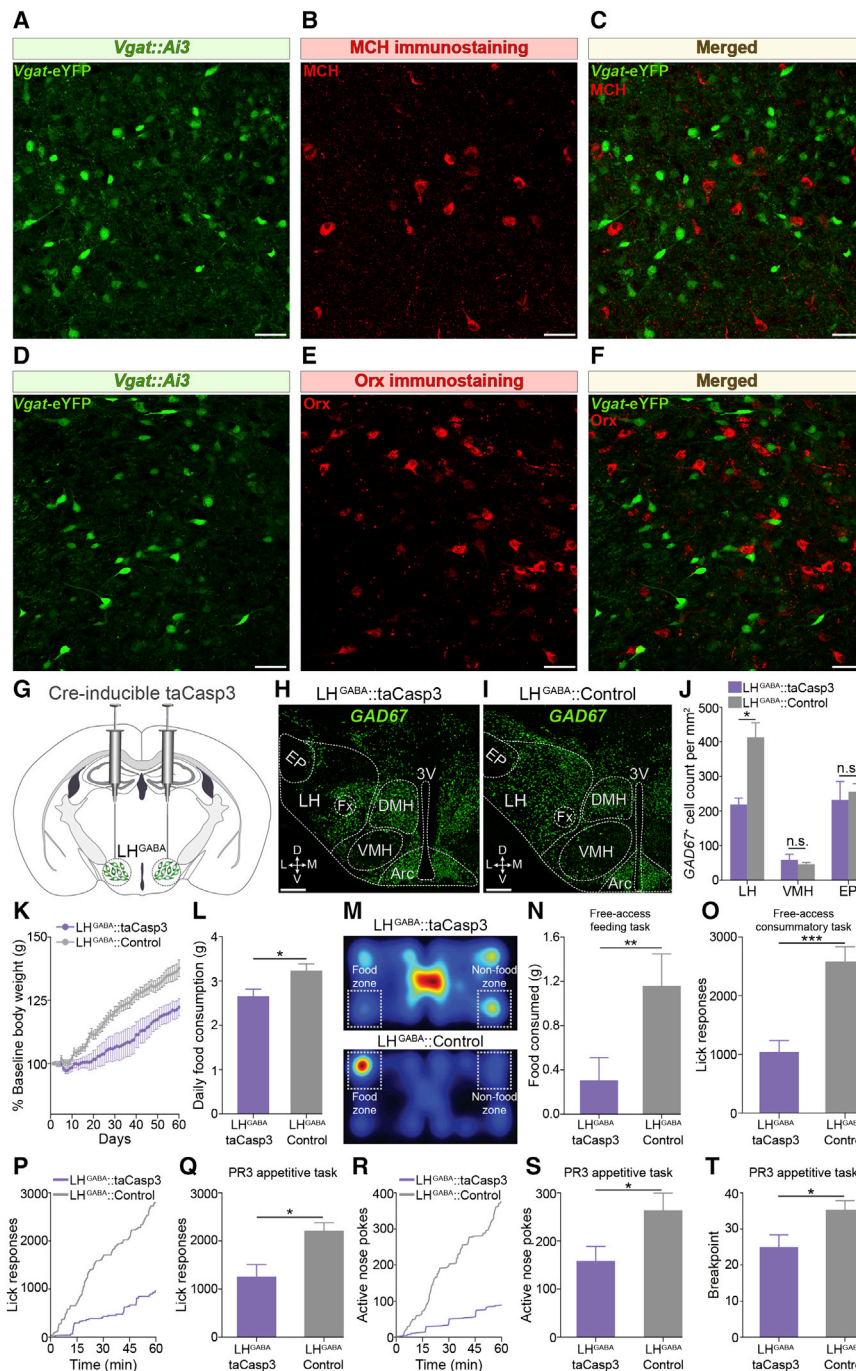


Figure 3. Genetic Ablation of LH *Vgat* Neurons that Are Separate from MCH and Orx Cells Attenuates Weight Gain, Food Seeking, and Consummatory Behaviors

(A–C) 20x confocal images of the dorsolateral LH demonstrating the absence of colocalization between *Vgat*-eYFP and MCH immunostaining ($n = 223 \pm 13.77$ *Vgat*-eYFP cells per mm², $n = 76 \pm 11.37$ MCH cells per mm², and 0% overlap, $n = 3$ LH sections; $n = 3$ *Vgat*-eYFP mice). Scale bars, 100 μ m.

(D–F) Representative 20x confocal images from a *Vgat*-eYFP brain slice (D) immunostained for Orx in red (E) displaying a lack of eYFP and Orx-immunoreactivity coexpression in *Vgat*⁺ LH cells (F); $n = 266 \pm 24.2$ *Vgat*-eYFP cells per mm², $n = 200 \pm 4.41$ Orx cells per mm², and 0% overlap; $n = 3$ LH sections; $n = 3$ *Vgat*-eYFP mice). Scale bars, 100 μ m.

(G) Viral injection of AAV2-FLEX-taCasp3-TEVp into the LH of *Vgat*-IRES-Cre mice.

(H and I) 20x confocal images demonstrating decreased *GAD67* expression in *LH^{GABA::}taCasp3* mice (H) compared to *LH^{GABA::}Control* animals (I). Scale bars, 200 μ m.

(J) *GAD67* expression is significantly decreased in the LH of *LH^{GABA::}taCasp3* animals compared to *LH^{GABA::}Controls*. Ablation of LH GABAergic neurons does not significantly alter *GAD67*-expression levels within the VMH and EP of *LH^{GABA::}taCasp3* and *LH^{GABA::}Control* mice ($n = 3$ LH sections; $n = 3$ mice per group; $F_{2,15} = 5.58$; $p = 0.01$).

(K) Ablation of LH GABAergic neurons significantly blunted weight gain induced from a calorie-dense diet ($n = 7$ mice per group; $F_{1,720} = 377.01$; $p = 0.0174$).

(L) Four weeks after taCasp3-TEVp viral injection, LH GABAergic cell death significantly reduced daily consumption of a calorie-dense diet ($n = 7$ mice per group; $t_{12} = 2.597$; $p = 0.0234$).

(M) Color map locations from example *LH^{GABA::}taCasp3* (top) and *LH^{GABA::}Control* (bottom) mice during the free-access feeding paradigm.

(N) *LH^{GABA::}taCasp3* mice display significant decreases in food consumption when compared to controls during the free-access feeding paradigm ($n = 7$ mice per group; $t_{12} = 3.239$; $p = 0.0071$).

(O) LH GABAergic neuron ablation significantly decreases lick responses in *LH^{GABA::}taCasp3* animals compared to controls during a free-access caloric consumption paradigm ($n = 7$ mice per group; $t_{12} = 5.3320$; $p = 0.0002$).

(P) Lick responses from example *LH^{GABA::}taCasp3* and *LH^{GABA::}Control* animals during a single (1 hr) PR3 session.

(Q) Ablation of LH GABAergic neurons significantly decreases lick responses in *LH^{GABA::}taCasp3* animals compared to *LH^{GABA::}Controls* during the PR3 task ($n = 6$ mice per group; $t_{10} = 3.024$; $p = 0.012$).

(R) Nose poke responses from example *LH^{GABA::}taCasp3* and *LH^{GABA::}Control* animals during a single PR3 session.

(S) Ablation of LH GABAergic neurons significantly decreases nose poke responding in *LH^{GABA::}taCasp3* animals compared to *LH^{GABA::}Controls* during the PR3 task ($n = 6$ mice per group; $t_{10} = 2.773$; $p = 0.019$).

(T) *LH^{GABA::}taCasp3* mice display significantly lower break points when compared to *LH^{GABA::}Controls* during the PR3 session ($n = 6$ mice per group; $t_{10} = 2.692$; $p = 0.022$).

All values are mean \pm SEM. Student's *t* test or two-way ANOVA. See also Figures S3 and S4.

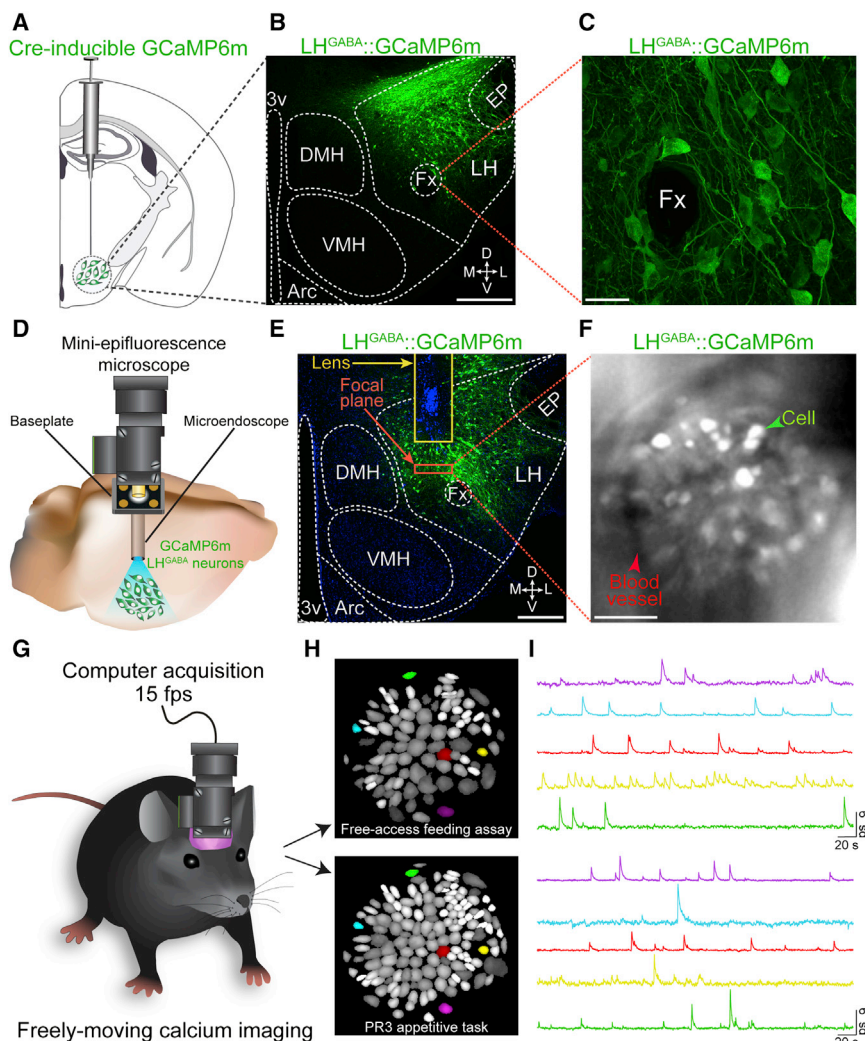


Figure 4. In Vivo Ca^{2+} Imaging of LH GABAergic Neurons in Freely Moving Mice

(A) Diagram showing unilateral viral injection of AAVDJ-EF1 α -DIO-GCaMP6m into the LH of *Vgat-IRES-Cre* mice.

(B) 10x confocal image of GCaMP6m expression in LH GABAergic neurons. Scale bar, 500 μm .

(C) 63x confocal image demonstrating stable and healthy GCaMP6m expression in LH GABAergic neurons several months after viral delivery. Scale bar, 20 μm .

(D) Integration of the miniepiilluminescence microscope with the microendoscope for deep-brain imaging of LH GABAergic neurons expressing GCaMP6m.

(E) 20x confocal image showing lens (microendoscope) placement and GCaMP6m-expressing neurons within the LH. Focal plane in tissue is 300 μm from the bottom of the lens as indicated by the red box. Scale bar, 500 μm .

(F) In vivo miniepiilluminescence image of GCaMP6m expression in the LH. Green arrow directs to a LH GABAergic neuron expressing GCaMP6m. Red arrow highlights a blood vessel. Scale bar, 100 μm .

(G) Illustration of the in vivo Ca^{2+} imaging setup.

(H) Schematized cell map of an example animal's LH GABAergic neurons visualized during free-access feeding and PR3 tasks. The same neurons can be tracked between sessions (colored cells).

(I) Ca^{2+} traces of individual neurons tracked in (H). See also Figure S5 and Movie S1.

characterized the response profiles of individual LH GABAergic neurons during the same free-access feeding task used above and recorded Ca^{2+} signals as food-restricted LH^{GABA::}GCaMP6m mice freely explored an arena that possessed discrete food-containing zones (FZ) in two of the outer corners (Figure 5A; Movie S2). This approach allowed for a visual representation of discrete spatial Ca^{2+} responses (Figures 5B and 5C) with some neurons exhibiting increased Ca^{2+} spiking in the presence of food (Figures 5B and 5D) and others showing decreased Ca^{2+} transients while the animal was in the FZ (Figures 5C and 5E). Further, we calculated a response ratio for each LH GABAergic neuron based on the frequency of Ca^{2+} responses in the FZ over the frequency in nonfood zones (FZ to NFZ; Figure 5F). We categorized response profiles from individual neurons as food zone excited (FZe) or food zone inhibited (FZI) if their log FZ/NFZ ratios were $>\pm 1$ SD of the mean (0.0 ± 0.3) of the entire population ($n = 612$ cells). Total Ca^{2+} activity from both groups decreased over time (Figure S6); however, average responses to FZ and NFZ areas within FZe and FZI neurons revealed significant differences in their respective directions (Fig-

ures 5G and 5H), supporting the design of our classification model. To spatially represent the response profiles of each neuron, we pseudocolored individual cells from an example animal's cell map by its log-transformed response ratio and observed that cells of differing response profiles intermingled with each other rather than segregating into separate clusters (Figure 5I). These data demonstrate that the endogenous activity dynamics of subsets of LH GABAergic neurons are preferentially modulated in environmental locations paired with food.

Individual LH GABAergic Neurons Selectively Encode Aspects of Appetitive or Consummatory Behaviors

Next, we sought to determine whether individual LH GABAergic neurons selectively compute aspects of appetitive and/or consummatory behaviors. Therefore, we tracked Ca^{2+} dynamics in LH GABAergic neurons during the same PR3 task described above, where LH^{GABA::}GCaMP6m mice worked to obtain a caloric liquid reward. Numerous individual LH GABAergic neurons showed time-locked Ca^{2+} transients in response to either the first lick after reward delivery (consummatory responses) or unreinforced active nose pokes (appetitive responses; Figures 6A and 6B; Movie S3). Nonconsummatory licks, those following unreinforced nose pokes, also evoked Ca^{2+} changes but to a much lesser degree (Figure S7A),

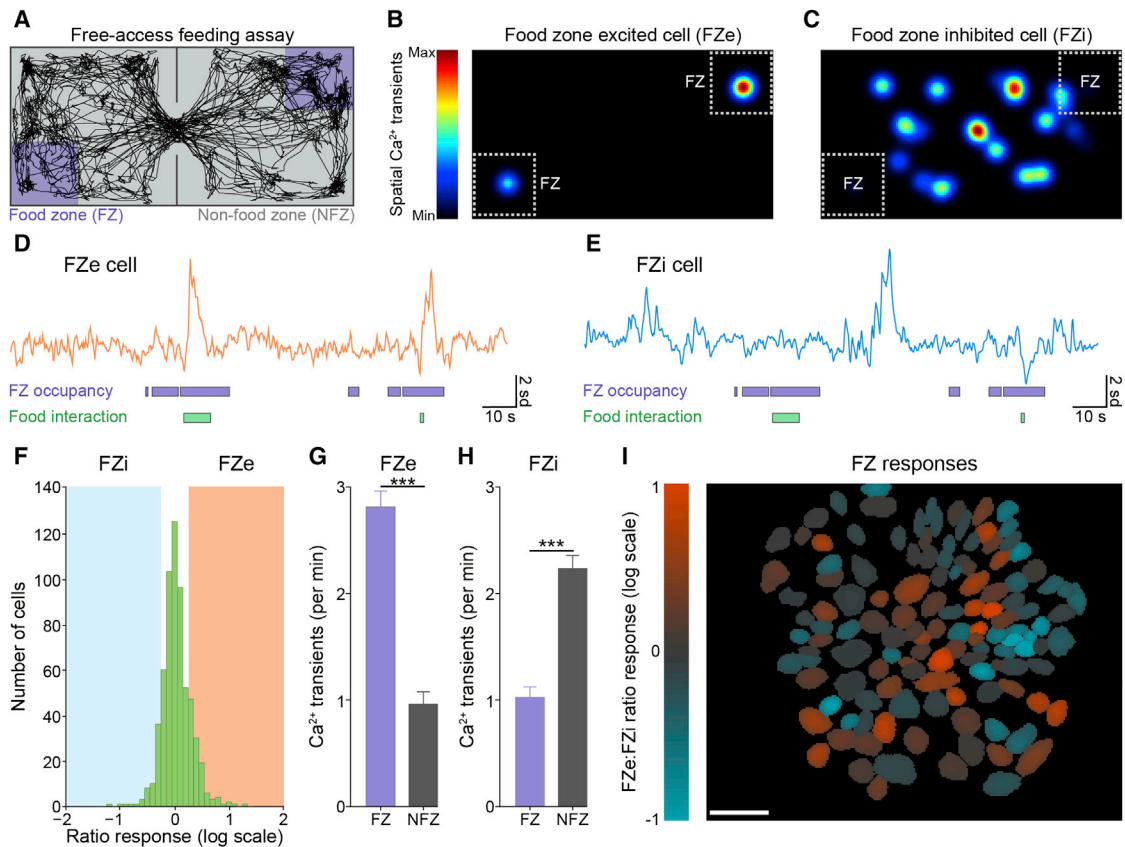


Figure 5. Subsets of LH GABAergic Neurons Display Enhanced or Reduced Activity to Environmental Locations Containing Food

(A) Example trace of animal's location during the free-access feeding task.

(B and C) Spatial Ca^{2+} activity maps of a single FZe (B) and FZI (C) cell. Behavioral arena is divided into 0.33×0.33 cm bins, where number of Ca^{2+} events per unit time is represented in color.

(D and E) Example Ca^{2+} traces from one FZe (orange; D) and one FZI (blue; E) cell during the free-access feeding task in relation to animal's location and state of interaction with food.

(F) Distribution of food zone (FZ) responses from all detected cells (mean = 0.0; SD = 0.3). FZe cells are classified as above one SD from the mean. FZI cells fall below one SD from the mean ($n = 612$ total cells imaged; $n = 87$ FZe cells; $n = 73$ FZI cells; $n = 6$ mice).

(G and H) Average Ca^{2+} transients per min for FZe (G) and FZI (H) cells in each zone ($n = 87$ FZe cells; $t_{86} = 14.92$; $p < 0.0001$; $n = 73$ FZI cells; $t_{72} = 15.08$, $p < 0.0001$).

(I) Example cell map with cells' color encoding response to FZ. Scale bar, 100 μm .

Error bars represent SEM. Student's t test. See also Figure S6 and Movie S2.

signifying that these consummatory Ca^{2+} responses depend on the presence of the caloric reward. We profiled all imaged neurons based on their differences between the average Ca^{2+} activity 1.5 s before and after each behavioral event to quantify their responses to aspects of feeding. LH GABAergic neurons displayed a variety of classifiable response profiles, such as cells that were modulated during reward consumption or immediately following active nose pokes (Figures 6C and 6D; Figures S7B–S7E). Thus, we categorized these neurons as responsive to a particular behavioral event (consumption or nose poke) if their Ca^{2+} activity (averaged from 0 to 1.5 s after the event) was statistically enhanced compared to the activity -1.5 to 0 s before the event. Averaging responses to consummatory licks revealed separate subpopulations of neurons with distinct Ca^{2+} responses (Figure 6E). A much larger population of LH GABAergic neurons displayed significant Ca^{2+} signals time-

locked to appetitive nose pokes (Figure 6F), although the amplitudes of these responses were lower compared to consummatory responses. LH GABAergic neurons displayed a diverse range of responses to consummatory licks and nose pokes (Figures S7F and S7G). However, we found that individual neurons that significantly responded to reward consumption were largely separate from those that respond to appetitive nose pokes (Figures 6G and 6H). Taken together, these data demonstrate that functionally segregated subpopulations of LH GABAergic neurons encode aspects of appetitive (nose-poke-responsive cells) or consummatory (lick-responsive cells) behaviors, but rarely both.

Lastly, we explored the Ca^{2+} dynamics of the same neurons between both behavioral paradigms (free-access feeding and the PR3 task) to determine if the same neurons that respond to the presence of food during the free-access feeding task also

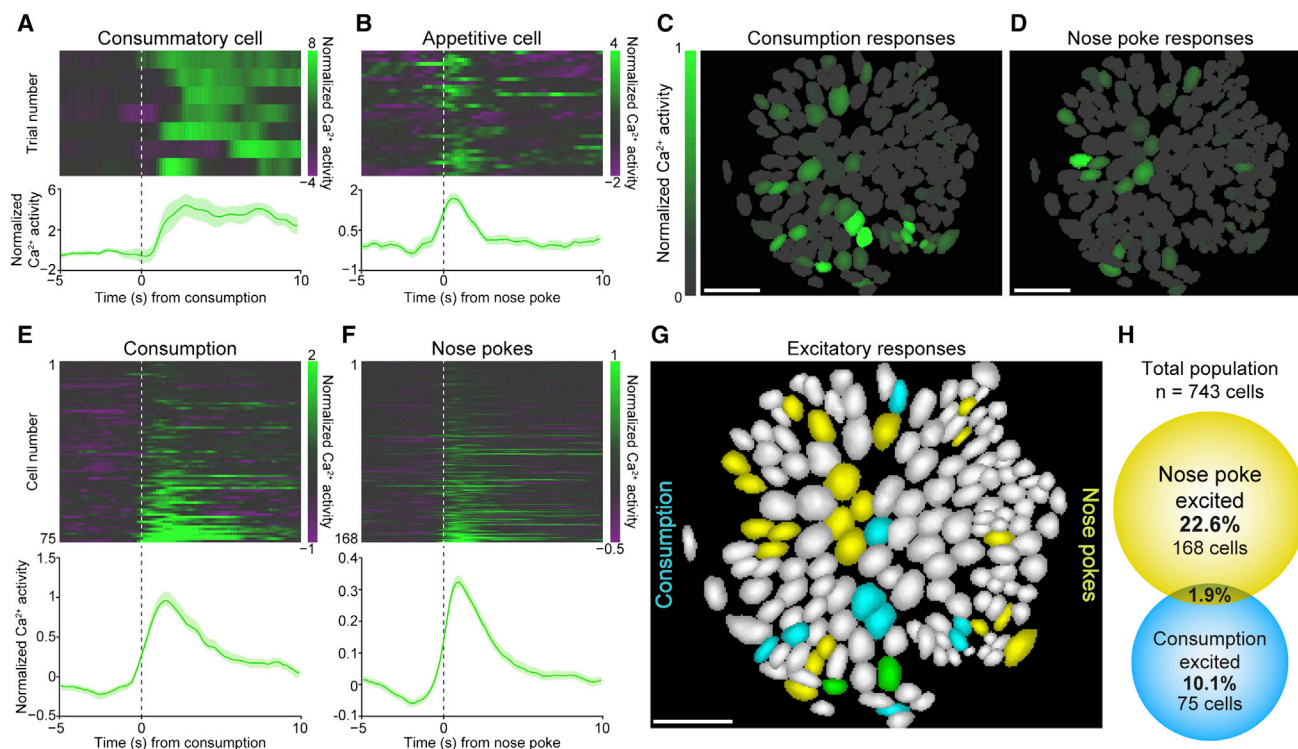


Figure 6. Separate LH GABAergic Neurons Selectively Encode Aspects of Appetitive or Consummatory Behaviors

(A) Ca^{2+} response to consummatory licks. (Top) Ca^{2+} response to individual consummatory licks from an example cell. (Bottom) Average Ca^{2+} response to all consummatory licks from the example cell.

(B) Ca^{2+} response to nose pokes. (Top) Ca^{2+} response to individual nose pokes from an example cell. (Bottom) Average Ca^{2+} response to all nose pokes from the example cell.

(C and D) Cell activity maps from an example animal. Color codes consummatory lick or appetitive nose-poke responses (average difference between Ca^{2+} signals from 1.5 s before and after the respective event) for each cell. Scale bars, 100 μm .

(E) Average Ca^{2+} response to consummatory licks following reward delivery from all excited cells from all animals. (Top) Average Ca^{2+} response to consummatory licks following reward delivery from all cells. (Bottom) Ca^{2+} response to consummatory licks averaged across cells excited by consummatory licks ($n = 75$ lick excited; $n = 743$ total cells).

(F) Average Ca^{2+} response to nose pokes from all excited cells of all animals. (Top) Average Ca^{2+} response to nose pokes from all cells. (Bottom) Ca^{2+} response to nose pokes averaged across cells excited by nose pokes ($n = 168$ nose poke excited; $n = 743$ total cells).

(G) Cell map from example animal. Cells excited by consummatory licks (cyan), nose pokes (yellow), or both (green). Scale bar, 100 μm .

(H) Venn diagram representing distribution and overlap of classified responsive cells in the PR3 task.

Green-shaded regions represent SEM. See also Figure S7 and Movie S3.

responded to consummatory licks and/or nose pokes during the PR3 task. To accomplish this, we registered IC unit cell maps from the same animal between sessions (Figures 7A–7C) and applied a 5 μm cutoff between center points of paired cells based on the distribution of cells from different imaging session (Figure 7D) and the distance between cells from the same sessions (Figure 7E). Using these criteria, only a subset of neurons imaged during the free-access feeding behavioral task were also detected in the PR3 task ($n = 40$ PR3 excited cells out of 125 FZ-responsive cells from paired sessions). Total paired cells between recording sessions: 472; Figure 7F). Neurons tracked in both sessions were not localized to any portion of the field of view and showed no distinct anatomical pattern or layout (Figures 7G–7I). However, paired cells displayed a high degree of functional overlap in the schematized IC unit cell maps between distinct imaging sessions. A large proportion of FZe neurons that

also respond to PR3 behavioral events (28/40 cells) displayed increased activity to either nose pokes or consummatory licks in the PR3 task, whereas a smaller portion of FZi neurons (12/40 cells) responded to PR3 behavioral events (Figure 7J). Taken together, these data reveal that subsets of LH GABAergic neurons functionally overlap between tasks that require different behavioral processes to obtain food, signifying the flexibility and complexity of these neurons for integrating and regulating components of feeding.

DISCUSSION

Historically, the LH has been viewed as a critical governor of both appetitive and consummatory behaviors (Hoebel and Teitelbaum, 1962; Margules and Olds, 1962). However, given that the LH encompasses a plethora of genetically, anatomically,

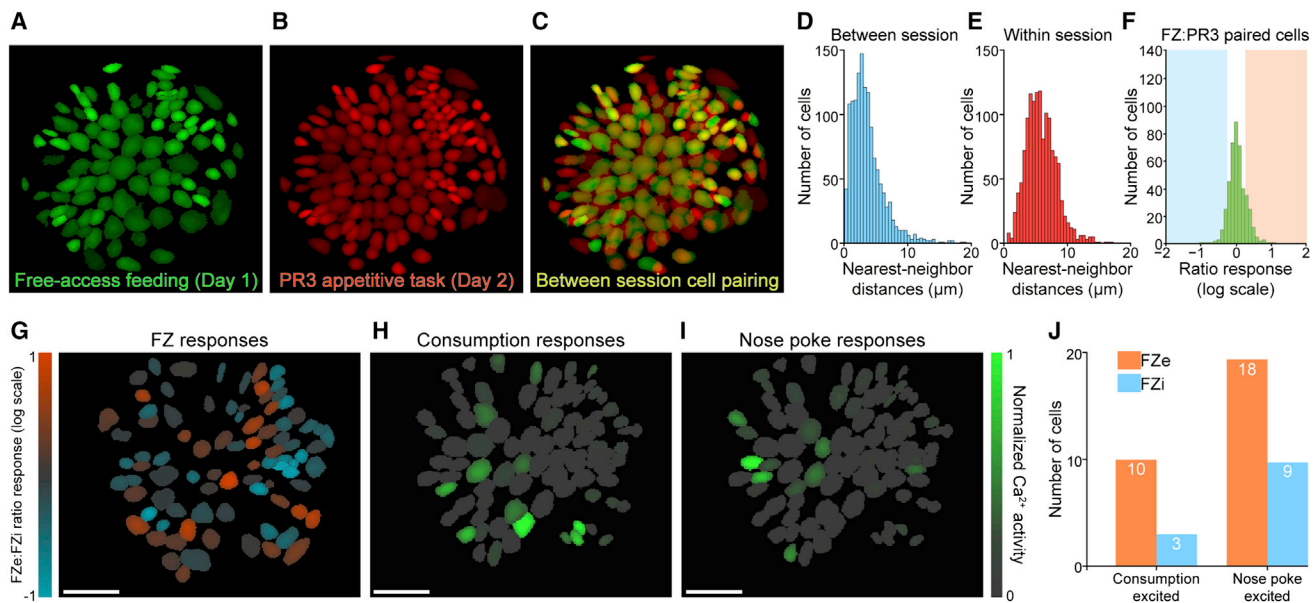


Figure 7. Tracking the Activity Dynamics of Individual LH GABAergic Neurons across Separate Behavioral Tasks

(A) Cell map from example animal during free-access feeding task.
 (B) Cell map from same example animal during PR3 appetitive task.
 (C) Merge of free-access feeding and PR3 appetitive task cell maps from same example animal.
 (D) Distribution of nearest-neighbor distances between cells of different behavioral tasks but within the same animal for all subjects.
 (E) Distribution of nearest-neighbor distances between all cells within the same behavioral task and imaging session.
 (F) Distribution of cell responses in free-access feeding task of only paired cells (n = 472 cells from 6 mice; n = 73 FZe cells; n = 52 FZI cells).
 (G–I) Maps of only paired cells from an example animal across the free-access feeding and PR3 tasks. Scale bars, 100 μm.
 (J) Bar graph showing cells that respond in both the free-access feeding and PR3 tasks.

and functionally distinct neurons that utilize various neurotransmitters and neuropeptides, the precise mechanisms by which these cell populations orchestrate behavior have remained a mystery. The results described here demonstrate that *Vgat*-expressing neurons in the LH function to promote both appetitive and consummatory behaviors. Importantly, we show that these *Vgat*-targeted cells in the LH do not colocalize with either of the neuropeptides, MCH, or Orx, in agreement with previous findings that certain LH GABAergic subpopulations (*GAD65*-expressing neurons) are neurochemically and electrophysiologically distinct from MCH and Orx cells (Karnani et al., 2013). However, it has also been reported that some MCH neurons do express GABAergic markers, such as *GAD67*, and release GABA from distal synaptic terminals (Jego et al., 2013). Thus, transgene penetrance in *Vgat-IRES-Cre* mice may be reduced compared to endogenous LH *Vgat* expression and/or MCH neurons may express extremely low levels of *Vgat*, which can be dynamically regulated as observed in other GABAergic neuronal populations (Jarvie and Hentges, 2012; Lamas et al., 2001; Sperk et al., 2003). In addition, the present results do not rule out the possibility that these unique *Vgat*-targeted GABAergic neurons contain other neuropeptides, such as Neuropeptide Y or Galanin (Allen and Cechetto, 1995; Goforth et al., 2014; Laque et al., 2013). Thus, parceling out precise neuronal subtypes from these multidimensional populations in the LH to then elucidate their function still remains a major challenge.

Divergent circuit connectivity between LH neurons and upstream and downstream circuit nodes could also account for the diverse appetitive and consummatory response profiles of the LH GABAergic neuronal population. Previous retrograde and anterograde tracing studies demonstrate strong connectivity between the LH and other feeding- and reward-associated brain regions that are also important for motivated behaviors, including other hypothalamic nuclei, midbrain, hindbrain, and striatal structures (Adamantidis et al., 2011; Betley et al., 2013; Gutierrez et al., 2011; Hahn and Swanson, 2012; Kempadoo et al., 2013; Leininger et al., 2011), implying that the selective-coding properties within separate LH GABAergic subpopulations might be input and/or projection dependent.

While the LH controls both appetitive and consummatory behaviors, these processes are encoded by distinct cellular LH GABAergic subpopulations. Our data suggest that separate subsets of appetitive-coding and consumption-coding ensembles exist within the LH GABAergic network. Thus, the LH GABAergic network can be viewed as a mosaic of functionally and computationally distinct cell types, requiring further definition. Nevertheless, these important computational differences among individual LH GABAergic neurons would have gone unnoticed if only bulk neuromodulatory approaches were employed, further underscoring the necessity of identifying the natural activity dynamics within a network to better understand the precise neural underpinnings of complex behavioral states.

EXPERIMENTAL PROCEDURES

Experimental Subjects

All procedures were conducted in accordance with the Guide for the Care and Use of Laboratory Animals, as adopted by the NIH, and with approval of the Institutional Animal Care and Use Committee at the University of North Carolina at Chapel Hill (UNC). Adult (25–30 g) male *Vgat-IRES-Cre* (Vong et al., 2011) or wild-type littermate mice were used.

Behavioral Assays

All food-deprived mice were restricted to 85% to 90% of their initial body weight by administering one daily feeding of ~2.5 to 3.0 g of standard grain-based chow immediately following each behavioral experiment, if performed. Animals were run on free-access feeding, real-time place preference, optical self-stimulation, and operant feeding (FR1 and PR3) assays. For optogenetic manipulations, light from solid-state lasers (473 or 532 nm) was delivered via custom-made patch cables to implanted chronic fibers on animals' head as described previously (Jennings et al., 2013b). For each chemogenetic behavioral manipulation, mice received i.p. injections of either vehicle or CNO (1 mg/kg) 45 min prior to testing with at least 3 days in between each session (counterbalanced). For further details, refer to the [Extended Experimental Procedures](#).

Freely Moving Ca^{2+} Imaging

A miniature microscope with an integrated LED (475 nm) was used to image GCaMP6m fluorescence in LH GABAergic neurons through an implanted microendoscope. Using nVista HD Acquisition Software (v. 2; Inscopix), images were acquired at 15 frames per second with the LED transmitting 0.1 to 0.2 mW of light on average. Ca^{2+} imaging was synchronized with time-stamped behavioral data at the start of each session. All images were processed using the Mosaic software (v. 1.0.0b; Inscopix) and then analyzed with custom MATLAB scripts. For further details, refer to the [Extended Experimental Procedures](#).

Statistical Analysis

Mean values are accompanied by SEM values. Comparisons were tested using paired or unpaired *t* tests. Two-way ANOVA tests followed by Bonferroni post hoc comparisons were applied for comparisons with more than two groups.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and three movies and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.12.026>.

AUTHOR CONTRIBUTIONS

J.H.J., R.L.U., and G.D.S. designed the experiments. J.H.J., R.L.U., S.L.R., A.M.S., J.G.T., J.H., K.V., P.A.K., M.A., and K.S.-S. conducted the experiments. C.R. and K.D. provided critical reagents. R.L.U., J.H.J., S.L.R., A.M.S., S.O., and G.D.S. analyzed the data. G.D.S., J.H.J., and R.L.U. wrote the manuscript.

ACKNOWLEDGMENTS

We thank Spencer Smith, Ilana Witten, Mark Schnitzer, and the members of the Stuber lab for helpful discussion. We thank the UNC vector core for viral packaging, Bradford Lowell and Linh Vong for the *Vgat-IRES-Cre* mice, and the UNC Neuroscience Center Microscopy Core (P30 NS045892). This work is dedicated to Johnathon Gregory Taylor (1996–2014), whose contributions were greatly appreciated and valued for this project. This study was supported by The Klarman Family Foundation, The Brain and Behavior Research Foundation, the Foundation for Prader-Willi Research, the Foundation of Hope, the National Institute on Drug Abuse (DA032750 and DA038168), and the National Institute on Alcohol Abuse and Alcoholism (P60 AA011605). J.H.J. was supported by a grant from the National Institute of Mental Health (MH104013), and A.M.S. was supported by grants from the National Institute on Drug Abuse

(DA034472). S.L.R. was supported by T32HD040127. S.O. is an employee of Inscopix, Inc.

Received: July 31, 2014

Revised: October 28, 2014

Accepted: November 24, 2014

Published: January 29, 2015

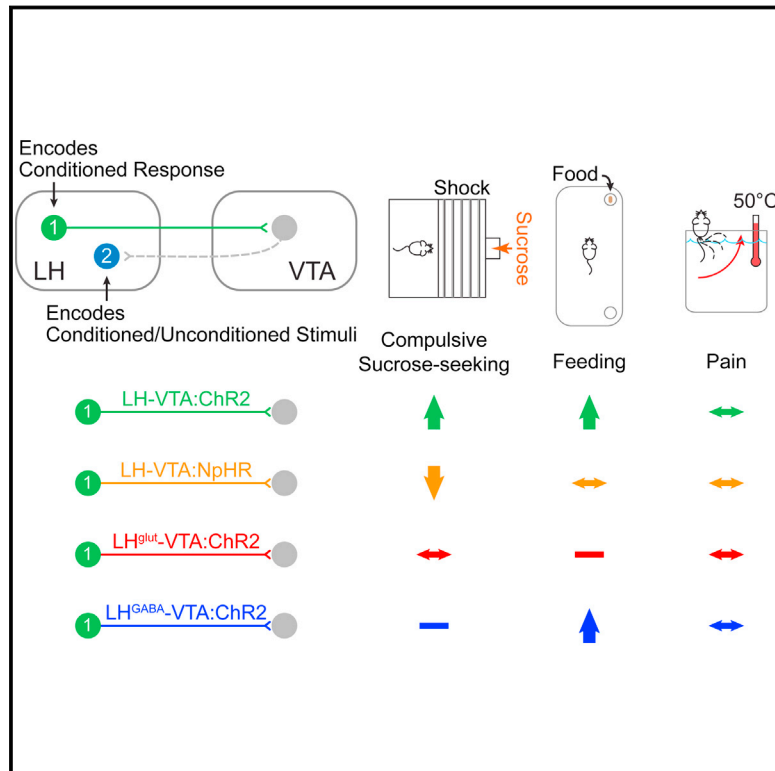
REFERENCES

- Adamantidis, A., and de Lecea, L. (2009). The hypocretins as sensors for metabolism and arousal. *J. Physiol.* 587, 33–40.
- Adamantidis, A.R., Tsai, H.-C., Boutrel, B., Zhang, F., Stuber, G.D., Budygin, E.A., Touriño, C., Bonci, A., Deisseroth, K., and de Lecea, L. (2011). Optogenetic interrogation of dopaminergic modulation of the multiple phases of reward-seeking behavior. *J. Neurosci.* 31, 10829–10835.
- Alexander, G.M., Rogan, S.C., Abbas, A.I., Armbruster, B.N., Pei, Y., Allen, J.A., Nonneman, R.J., Hartmann, J., Moy, S.S., Nicolelis, M.A., et al. (2009). Remote control of neuronal activity in transgenic mice expressing evolved G protein-coupled receptors. *Neuron* 63, 27–39.
- Allen, G.V., and Cechetto, D.F. (1995). Neurotensin in the lateral hypothalamic area: origin and function. *Neuroscience* 69, 533–544.
- Anand, B.K., and Brobeck, J.R. (1951). Hypothalamic control of food intake in rats and cats. *Yale J. Biol. Med.* 24, 123–140.
- Ball, G.F., and Balthazart, J. (2008). How useful is the appetitive and consummatory distinction for our understanding of the neuroendocrine control of sexual behavior? *Horm. Behav.* 53, 307–311, author reply 315–318.
- Barretto, R.P.J., Ko, T.H., Jung, J.C., Wang, T.J., Capps, G., Waters, A.C., Ziv, Y., Attardo, A., Recht, L., and Schnitzer, M.J. (2011). Time-lapse imaging of disease progression in deep brain areas using fluorescence microendoscopy. *Nat. Med.* 17, 223–228.
- Berthoud, H.-R., and Münzberg, H. (2011). The lateral hypothalamus as integrator of metabolic and environmental needs: from electrical self-stimulation to opto-genetics. *Physiol. Behav.* 104, 29–39.
- Betley, J.N., Cao, Z.F.H., Ritola, K.D., and Sternson, S.M. (2013). Parallel, redundant circuit organization for homeostatic control of feeding behavior. *Cell* 155, 1337–1350.
- Chow, B.Y., Han, X., Dobry, A.S., Qian, X., Chuong, A.S., Li, M., Henninger, M.A., Belfort, G.M., Lin, Y., Monahan, P.E., and Boyden, E.S. (2010). High-performance genetically targetable optical neural silencing by light-driven proton pumps. *Nature* 463, 98–102.
- Craig, W. (1917). Appetites and aversions as constituents of instincts. *Proc. Natl. Acad. Sci. USA* 3, 685–688.
- Elias, C.F., Saper, C.B., Maratos-Flier, E., Tritos, N.A., Lee, C., Kelly, J., Tatro, J.B., Hoffman, G.E., Ollmann, M.M., Barsh, G.S., et al. (1998). Chemically defined projections linking the mediobasal hypothalamus and the lateral hypothalamic area. *J. Comp. Neurol.* 402, 442–459.
- Erlander, M.G., Tillakaratne, N.J., Feldblum, S., Patel, N., and Tobin, A.J. (1991). Two genes encode distinct glutamate decarboxylases. *Neuron* 7, 91–100.
- Flusberg, B.A., Nimmerjahn, A., Cocker, E.D., Mukamel, E.A., Barretto, R.P.J., Ko, T.H., Burns, L.D., Jung, J.C., and Schnitzer, M.J. (2008). High-speed, miniaturized fluorescence microscopy in freely moving mice. *Nat. Methods* 5, 935–938.
- Ghosh, K.K., Burns, L.D., Cocker, E.D., Nimmerjahn, A., Ziv, Y., Gamal, A.E., and Schnitzer, M.J. (2011). Miniaturized integration of a fluorescence microscope. *Nat. Methods* 8, 871–878.
- Goforth, P.B., Leininger, G.M., Patterson, C.M., Satin, L.S., and Myers, M.G., Jr. (2014). Leptin acts via lateral hypothalamic area neurotensin neurons to inhibit orexin neurons by multiple GABA-independent mechanisms. *J. Neurosci.* 34, 11405–11415.
- Gutierrez, R., Lobo, M.K., Zhang, F., and de Lecea, L. (2011). Neural integration of reward, arousal, and feeding: recruitment of VTA, lateral hypothalamus, and ventral striatal neurons. *IUBMB Life* 63, 824–830.

- Hahn, J.D., and Swanson, L.W. (2012). Connections of the lateral hypothalamic area juxtadorsomedial region in the male rat. *J. Comp. Neurol.* 520, 1831–1890.
- Harrell, L.E., Decastro, J.M., and Balagura, S. (1975). A critical evaluation of body weight loss following lateral hypothalamic lesions. *Physiol. Behav.* 15, 133–136.
- Hoebel, B.G. (1965). Hypothalamic lesions by electrocauterization: disinhibition of feeding and self-stimulation. *Science* 149, 452–453.
- Hoebel, B.G., and Teitelbaum, P. (1962). Hypothalamic control of feeding and self-stimulation. *Science* 135, 375–377.
- Hökfelt, T., Broberger, C., Xu, Z.Q., Sergeev, V., Ubink, R., and Diez, M. (2000). Neuropeptides—an overview. *Neuropharmacology* 39, 1337–1356.
- Jarvie, B.C., and Hentges, S.T. (2012). Expression of GABAergic and glutamatergic phenotypic markers in hypothalamic proopiomelanocortin neurons. *J. Comp. Neurol.* 520, 3863–3876.
- Jego, S., Glasgow, S.D., Herrera, C.G., Ekstrand, M., Reed, S.J., Boyce, R., Friedman, J., Burdakov, D., and Adamantidis, A.R. (2013). Optogenetic identification of a rapid eye movement sleep modulatory circuit in the hypothalamus. *Nat. Neurosci.* 16, 1637–1643.
- Jennings, J.H., and Stuber, G.D. (2014). Tools for resolving functional activity and connectivity within intact neural circuits. *Curr. Biol.* 24, R41–R50.
- Jennings, J.H., Rizzi, G., Stamatakis, A.M., Ung, R.L., and Stuber, G.D. (2013a). The inhibitory circuit architecture of the lateral hypothalamus orchestrates feeding. *Science* 341, 1517–1521.
- Jennings, J.H., Sparta, D.R., Stamatakis, A.M., Ung, R.L., Pleil, K.E., Kash, T.L., and Stuber, G.D. (2013b). Distinct extended amygdala circuits for divergent motivational states. *Nature* 496, 224–228.
- Karnani, M.M., Apergis-Schoute, J., Adamantidis, A., Jensen, L.T., de Lecea, L., Fugger, L., and Burdakov, D. (2011). Activation of central orexin/hypocretin neurons by dietary amino acids. *Neuron* 72, 616–629.
- Karnani, M.M., Szabó, G., Erdélyi, F., and Burdakov, D. (2013). Lateral hypothalamic GAD65 neurons are spontaneously firing and distinct from orexin and melanin-concentrating hormone neurons. *J. Physiol.* 591, 933–953.
- Kempadoo, K.A., Tourino, C., Cho, S.L., Magnani, F., Leininger, G.-M., Stuber, G.D., Zhang, F., Myers, M.G., Deisseroth, K., de Lecea, L., and Bonci, A. (2013). Hypothalamic neurotensin projections promote reward by enhancing glutamate transmission in the VTA. *J. Neurosci.* 33, 7618–7626.
- Kimura, H., and Kuriyama, K. (1975). Distribution of gamma-aminobutyric acid (GABA) in the rat hypothalamus: functional correlates of GABA with activities of appetite controlling mechanisms. *J. Neurochem.* 24, 903–907.
- Knight, Z.A., Tan, K., Birsoy, K., Schmidt, S., Garrison, J.L., Wysocki, R.W., Emiliano, A., Ekstrand, M.I., and Friedman, J.M. (2012). Molecular profiling of activated neurons by phosphorylated ribosome capture. *Cell* 151, 1126–1137.
- Krashes, M.J., Koda, S., Ye, C., Rogan, S.C., Adams, A.C., Cusher, D.S., Maratos-Flier, E., Roth, B.L., and Lowell, B.B. (2011). Rapid, reversible activation of AgRP neurons drives feeding behavior in mice. *J. Clin. Invest.* 121, 1424–1428.
- Lamas, M., Gómez-Lira, G., and Gutiérrez, R. (2001). Vesicular GABA transporter mRNA expression in the dentate gyrus and in mossy fiber synaptosomes. *Brain Res. Mol. Brain Res.* 93, 209–214.
- Laque, A., Zhang, Y., Gettys, S., Nguyen, T.-A., Bui, K., Morrison, C.D., and Münzberg, H. (2013). Leptin receptor neurons in the mouse hypothalamus are colocalized with the neuropeptide galanin and mediate anorexigenic leptin action. *Am. J. Physiol. Endocrinol. Metab.* 304, E999–E1011.
- Lein, E.S., Hawrylycz, M.J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A.F., Boguski, M.S., Brockway, K.S., Byrnes, E.J., et al. (2007). Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445, 168–176.
- Leininger, G.M., Opland, D.M., Jo, Y.-H., Faouzi, M., Christensen, L., Cappelucci, L.A., Rhodes, C.J., Gnagy, M.E., Becker, J.B., Pothos, E.N., et al. (2011). Leptin action via neurotensin neurons controls orexin, the mesolimbic dopamine system and energy balance. *Cell Metab.* 14, 313–323.
- Lorenz, K.Z. (1950). The comparative method in studying innate behavior patterns. In *Physiological Mechanisms in Animal Behavior*. Society's Symposium IV (Oxford: Academic Press), pp. 221–268.
- Madisen, L., Zwingman, T.A., Sunken, S.M., Oh, S.W., Zariwala, H.A., Gu, H., Ng, L.L., Palmiter, R.D., Hawrylycz, M.J., Jones, A.R., et al. (2010). A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* 13, 133–140.
- Margules, D.L., and Olds, J. (1962). Identical “feeding” and “rewarding” systems in the lateral hypothalamus of rats. *Science* 135, 374–375.
- Mattis, J., Tye, K.M., Ferenczi, E.A., Ramakrishnan, C., O’Shea, D.J., Prakash, R., Gunaydin, L.A., Hyun, M., Fenno, L.E., Gradinaru, V., et al. (2012). Principles for applying optogenetic tools derived from direct comparative analysis of microbial opsins. *Nat. Methods* 9, 159–172.
- Molina-Borja, M., and Gómez-Soutullo, T. (1989). Electrical stimulation of lateral hypothalamic area and behavioural sequences in a lacertid lizard. *Behav. Brain Res.* 32, 197–201.
- Mukamel, E.A., Nimmerjahn, A., and Schnitzer, M.J. (2009). Automated analysis of cellular signals from large-scale calcium imaging data. *Neuron* 63, 747–760.
- Olds, J., and Milner, P. (1954). Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *J. Comp. Physiol. Psychol.* 47, 419–427.
- Ono, T., Nishino, H., Sasaki, K., Fukuda, M., and Muramoto, K. (1981). Long-term lateral hypothalamic single unit analysis and feeding behavior in freely moving rats. *Neurosci. Lett.* 26, 79–83.
- Ono, T., Sasaki, K., Nishino, H., Fukuda, M., and Shibata, R. (1986). Feeding and diurnal related activity of lateral hypothalamic neurons in freely behaving rats. *Brain Res.* 373, 92–102.
- Quaade, F., Vaernet, K., and Larsson, S. (1974). Stereotaxic stimulation and electrocoagulation of the lateral hypothalamus in obese humans. *Acta Neurochir. (Wien)* 30, 111–117.
- Sakurai, T., Amemiya, A., Ishii, M., Matsuzaki, I., Chemelli, R.M., Tanaka, H., Williams, S.C., Richardson, J.A., Kozlowski, G.P., Wilson, S., et al. (1998). Orexins and orexin receptors: a family of hypothalamic neuropeptides and G protein-coupled receptors that regulate feeding behavior. *Cell* 92, 573–585.
- Schallert, T., and Whishaw, I.Q. (1978). Two types of aphagia and two types of sensorimotor impairment after lateral hypothalamic lesions: observations in normal weight, dieted, and fattened rats. *J. Comp. Physiol. Psychol.* 92, 720–741.
- Sherrington, C.S. (1906). *The Integrative Action of the Nervous System* (Cambridge: CUP Archive).
- Sperk, G., Schwarzer, C., Heilman, J., Furtinger, S., Reimer, R.J., Edwards, R.H., and Nelson, N. (2003). Expression of plasma membrane GABA transporters but not of the vesicular GABA transporter in dentate granule cells after kainic acid seizures. *Hippocampus* 13, 806–815.
- Tinbergen, N. (1951). *The Study of Instinct* (New York: Clarendon Press/Oxford University Press).
- Vong, L., Ye, C., Yang, Z., Choi, B., Chua, S., Jr., and Lowell, B.B. (2011). Leptin action on GABAergic neurons prevents obesity and reduces inhibitory tone to POMC neurons. *Neuron* 71, 142–154.
- Whiddon, B.B., and Palmiter, R.D. (2013). Ablation of neurons expressing melanin-concentrating hormone (MCH) in adult mice improves glucose tolerance independent of MCH signaling. *J. Neurosci.* 33, 2009–2016.
- Wise, R.A. (1971). Individual differences in effects of hypothalamic stimulation: the role of stimulation locus. *Physiol. Behav.* 6, 569–572.
- Yang, C.F., Chiang, M.C., Gray, D.C., Prabhakaran, M., Alvarado, M., Juntti, S.A., Unger, E.K., Wells, J.A., and Shah, N.M. (2013). Sexually dimorphic neurons in the ventromedial hypothalamus govern mating in both sexes and aggression in males. *Cell* 153, 896–909.
- Ziv, Y., Burns, L.D., Cocker, E.D., Hamel, E.O., Ghosh, K.K., Kitch, L.J., El Gamal, A., and Schnitzer, M.J. (2013). Long-term dynamics of CA1 hippocampal place codes. *Nat. Neurosci.* 16, 264–266.

Decoding Neural Circuits that Control Compulsive Sucrose Seeking

Graphical Abstract



Authors

Edward H. Nieh, Gillian A. Matthews, ..., Craig P. Wildes, Kay M. Tye

Correspondence

kaytye@mit.edu

In Brief

A neural circuit loop between the lateral hypothalamus and the ventral tegmental area that selectively controls compulsive sugar consumption without preventing feeding necessary for survival provides a potential target for compulsive-overeating therapeutic interventions.

Highlights

- LH-VTA neurons encode reward-seeking actions after they transition to habits
- A subset of LH neurons downstream of VTA encode reward expectation
- LH-VTA projections provide bidirectional control over compulsive sucrose seeking
- Activating LH-VTA GABAergic projections increases maladaptive gnawing behavior



Decoding Neural Circuits that Control Compulsive Sucrose Seeking

Edward H. Nieh,^{1,2} Gillian A. Matthews,^{1,2} Stephen A. Allsop,^{1,2} Kara N. Presbrey,¹ Christopher A. Leppla,¹ Romy Wichmann,¹ Rachael Neve,¹ Craig P. Wildes,¹ and Kay M. Tye^{1,*}

¹The Picower Institute for Learning and Memory, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

²Co-first author

*Correspondence: kaytye@mit.edu

<http://dx.doi.org/10.1016/j.cell.2015.01.003>

SUMMARY

The lateral hypothalamic (LH) projection to the ventral tegmental area (VTA) has been linked to reward processing, but the computations within the LH-VTA loop that give rise to specific aspects of behavior have been difficult to isolate. We show that LH-VTA neurons encode the learned action of seeking a reward, independent of reward availability. In contrast, LH neurons downstream of VTA encode reward-predictive cues and unexpected reward omission. We show that inhibiting the LH-VTA pathway reduces “compulsive” sucrose seeking but not food consumption in hungry mice. We reveal that the LH sends excitatory and inhibitory input onto VTA dopamine (DA) and GABA neurons, and that the GABAergic projection drives feeding-related behavior. Our study overlays information about the type, function, and connectivity of LH neurons and identifies a neural circuit that selectively controls compulsive sugar consumption, without preventing feeding necessary for survival, providing a potential target for therapeutic interventions for compulsive-overeating disorder.

INTRODUCTION

Tremendous heterogeneity exists across lateral hypothalamic (LH) neurons in terms of function and connectivity, and this can be observed by the variety of behaviors related to reward, motivation, and feeding linked with this region. However, little is known about how the LH computes specific aspects of reward processing and how this information is relayed to downstream targets. Electrical stimulation of the LH produces intracranial self-stimulation (ICSS) (Olds and Milner, 1954), as well as grooming, sexual, and gnawing behaviors (Singh et al., 1996). LH neurons encode sensory stimuli (Norgren, 1970; Yamamoto et al., 1989), including reward-associated cues (Nakamura et al., 1987). LH neurons also fire during both feeding (Burton et al., 1976; Schwartzbaum, 1988) and drinking (Tabuchi et al., 2002). However, making sense of the remarkable functional heterogeneity observed in the LH has been a major challenge in the field.

Although the LH is interconnected with many subcortical regions, we have a poor understanding of how the functional and cellular heterogeneity of the LH is transposed upon these anatomical connections. One LH projection target of interest is the ventral tegmental area (VTA), a critical component in reward processing (Wise, 2004). The LH-VTA projection was explored in early studies that used electrophysiological recordings combined with antidromic stimulation (Bielajew and Shizgal, 1986; Gratton and Wise, 1988). It has since been confirmed, using a rabies-virus-mediated tracing approach, that there is monosynaptic input from LH neurons onto dopamine (DA) neurons in the VTA (Watabe-Uchida et al., 2012). The VTA also sends reciprocal projections back to the LH, both directly and indirectly via other regions such as the nucleus accumbens, amygdala, hippocampus, and ventral pallidum (Barone et al., 1981; Beckstead et al., 1979; Simon et al., 1979).

Although both electrical (Bielajew and Shizgal, 1986) and optical (Kempadoo et al., 2013) stimulation have established a causal role for the LH projection to the VTA in ICSS, several questions remain to be answered. First, what is the neural response of LH-VTA neurons to different aspects of reward-related behaviors? Second, what is the role of the LH-VTA projection in reward seeking under different reinforcement contingencies? Third, what is the overall composition of fast transmission mediated by LH inputs to the VTA, and which VTA cells receive excitatory/inhibitory input? Finally, what do the excitatory and inhibitory components of the LH-VTA pathway each contribute toward orchestrating the pursuit of appetitive reward?

To address these questions, we recorded from LH neurons in freely moving mice and used optogenetic-mediated photoidentification to overlay information about the naturally occurring neural computations during reward processing upon information about the connectivity of LH neurons. In addition, we used ex vivo patch-clamp experiments to explore the composition of GABAergic and glutamatergic LH inputs onto both DA and GABA neurons within the VTA. Building on our results from the recordings experiments, we utilized behavioral tasks to establish causal relationships between aspects of both reward seeking and feeding and the activation of distinct subsets of LH-VTA projections. Together, these data help us establish a model for how the components within the LH-VTA loop work together to process reward and how manipulating individual components can have profound effects on behavior.

RESULTS

Photoidentification of Distinct Components in the LH-VTA Circuit

In order to identify LH neurons that provide monosynaptic input to the VTA *in vivo* and observe their activity during freely moving behaviors, we used a dual-virus strategy to selectively express channelrhodopsin-2 (ChR2) in LH neurons providing monosynaptic input to the VTA (Figures 1A and S1). We injected an adeno-associated viral vector (AAV₅) carrying ChR2-eYFP in a Cre-recombinase-dependent double-inverted open reading frame (DIO) construct into the LH to infect local somata and injected a retrogradely traveling herpes simplex virus (HSV) carrying Cre-recombinase into the VTA. Subsequent recombination permitted opsin and fluorophore expression selectively in LH neurons providing monosynaptic input to the VTA. To confirm our approach, we performed *ex vivo* whole-cell patch-clamp recordings in horizontal brain slices containing the LH and recorded from neurons expressing ChR2-eYFP, as well as neighboring LH neurons that were ChR2-eYFP negative (Figure 1B). Light-evoked spike latencies, measured from light-pulse onset to the peak of the action potential, ranged from 3–8 ms (Figure 1C). We also found that none of the non-expressing (ChR2-negative) cells recorded showed excitatory responses to photostimulation ($n = 14$; Figure 1C), despite their proximity to ChR2-expressing cells.

In order to perform optogenetically mediated photoidentification *in vivo*, an optrode was implanted into the LH to record neuronal activity during a sucrose-seeking task. In the same recording session, we provided several patterns of photostimulation to identify ChR2-expressing LH-VTA neurons (Figures 1D and S1). We examined the distribution of excitatory photoresponse latencies across all LH neurons displaying a time-locked change in firing rate in response to illumination and observed a bimodal distribution (Figure 1E). We observed a population of neurons during *in vivo* recordings with latencies in a range of 3–8 ms. This was identical to the latency range found in ChR2-expressing LH-VTA neurons when we recorded *ex vivo*. We termed these units “Type 1” units (Figures 1C, 1E, and 1F). In addition, there was a distinct population of cells with ~100 ms photoresponse latencies (Figures 1E and 1G), and we termed these “Type 2” units. We also observed neurons that were inhibited in response to photostimulation of LH-VTA neurons (Figure S2), and we termed these “Type 3” units. We compared the action potential duration (as measured from peak to trough) and mean firing rates of Type 1 and Type 2 units as well as those that did not show a photoresponse (Figure 1H). The distribution of action potential durations of Type 1 (Figure 1I) and Type 2 (Figure 1J) units shows that the majority of Type 1 units have an action potential duration less than 500 μ s (84%; $n = 16/19$, binomial distribution, $p = 0.002$).

Although Type 1 units fit standard criteria to be classified as ChR2 expressing (Cohen et al., 2012; Zhang et al., 2013), it was unclear whether the longer latency photoresponse of Type 2 units was indicative of ChR2-expressing neurons that responded more slowly to photostimulation, or whether this effect was due to network activity. Given that the ChR2-expressing (Type 1) LH neurons project directly to the VTA, one possibility

was that Type 2 neurons were receiving feedback from the VTA (Figure 1K). Another possibility was that Type 2 neurons were activated by axon collaterals from Type 1 neurons (Figure 1L). To differentiate between these two possible circuit models, we inhibited the VTA in conjunction with photoidentification in the LH.

Long Latency Photoresponses in LH Neurons Are Mediated by Feedback from the VTA

Based on our circuit models, we would expect distal inhibition to have no effect on the photoresponses of ChR2-expressing LH neurons. However, if photoresponsive, but non-expressing, LH neurons relied on feedback from the VTA to elicit a time-locked response to illumination (Figure 1K), we would expect an attenuation of photoresponses in these neurons upon VTA inhibition. We expressed ChR2 in LH-VTA cells as above, but this time also expressed enhanced halorhodopsin 3.0 (NpHR) in the VTA and implanted an optic fiber in the VTA in addition to the optrode in LH (Figure 2A). We delivered the same blue-light illumination patterns in the LH for all three epochs but also photoinhibited the VTA with yellow light in the second epoch (Figure 2A).

The photoresponses of Type 1 units to blue-light illumination in the LH were unaffected by photoinhibition of the VTA, which is consistent with ChR2 expression in Type 1 LH-VTA neurons (Figure 2B). In contrast, the majority of Type 2 units (87%; $n = 13/15$, binomial distribution, $p = 0.004$) showed a significant attenuation of photoresponses to blue-light pulses delivered in the LH upon photoinhibition of VTA neurons. The responses of Type 1 and Type 2 units during VTA photoinhibition were significantly different (chi-square = 7.64, $p = 0.0057$; Figures 2B and 2C). These differences can also be seen in the max Z scores during individual epochs (Figure 2D) and with the yellow-ON epoch normalized to the yellow-OFF epoch (Figure 2E). These data suggest that Type 2 LH neurons receive input (either directly or indirectly) from the VTA (Figure 1K) rather than via local axon collaterals (Figure 1L).

Distinct Encoding Properties of LH Neurons Either upstream or downstream of the VTA

Having identified these two distinct types of LH neurons in the LH-VTA loop, we wanted to examine naturally occurring neural activity during a sucrose self-administration task (Figure 3A). Mice were trained to perform nosepoke responses for a cue predicting sucrose delivery at an adjacent port (as in Tye et al., 2008). To allow us to differentiate neural responses to the nosepoke and the cue, the cue and sucrose were delivered on a partial reinforcement schedule, wherein 50% of nosepokes were paired with a cue and sucrose delivery.

Type 1 units showed phasic responses to sucrose port entry, as seen in a representative Type 1 unit (Figure 3B), as well as the population data for all Type 1 units (Figure 3C). The phasic responses of Type 2 units, however, mainly reflected responses to the reward-predictive cue (Figures 3D and 3E). The normalized firing patterns of all recorded neurons ($n = 198$, divided into Type 1, 2, 3, and non-responsive units) are displayed for each task component: nosepokes paired with the cue, nosepokes in the absence of the cue, and sucrose port entry (Figure 3F). All Type 1 units that showed task-relevant phasic changes in activity

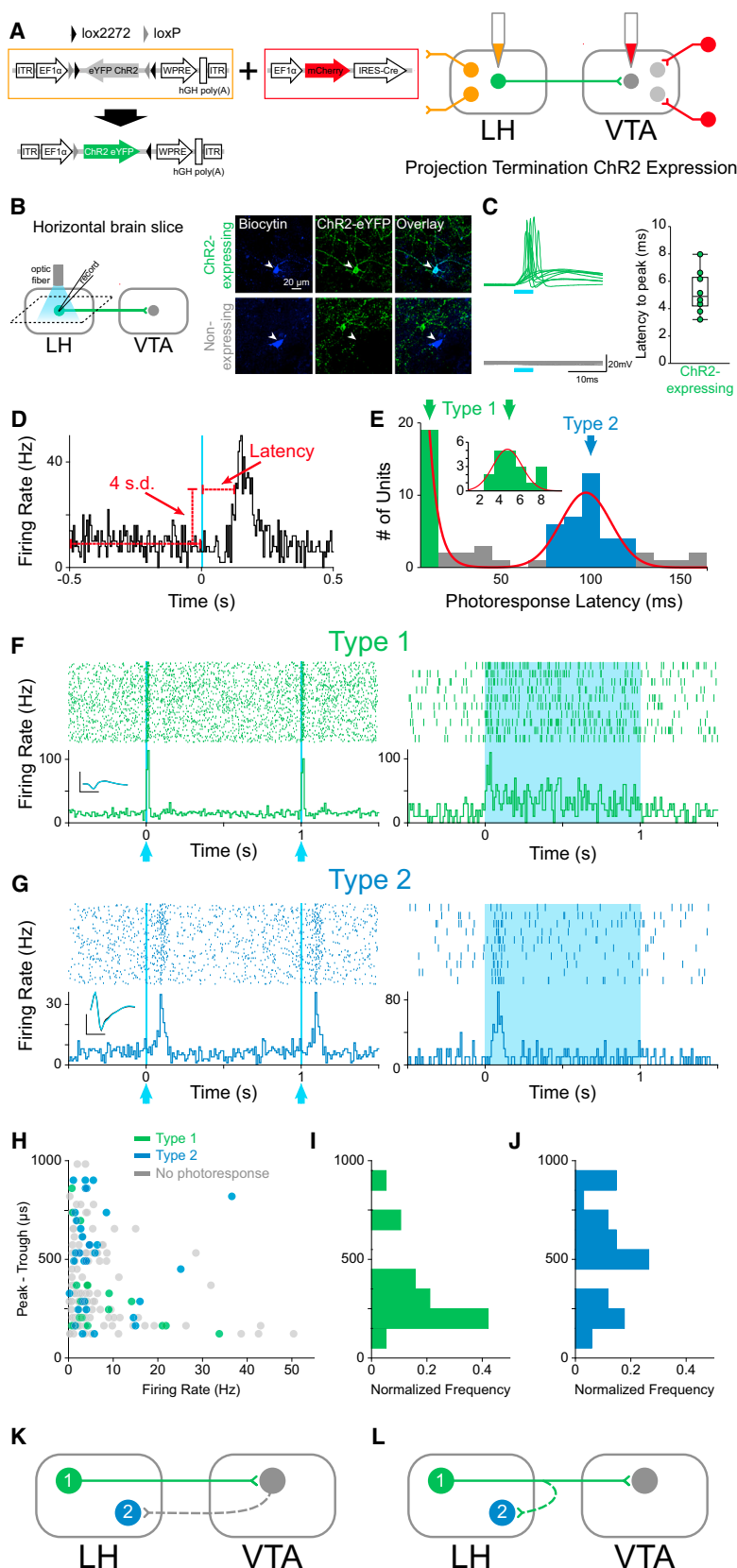


Figure 1. Phototagging LH-VTA Projections Reveals Two Populations of Neurons with Different Response Latencies to Photostimulation

(A) Wild-type mice ($n = 12$) were injected with AAV₅-DIO-ChR2-eYFP into the lateral hypothalamus (LH) and HSV-EF1 α -IRES-Cre-mCherry into the ventral tegmental area (VTA).

(B) Horizontal brain slices containing the LH were prepared for whole-cell patch-clamp recordings in ChR2-expressing and non-expressing LH neurons.

(C) Individual traces recorded in current-clamp mode showing the response of ChR2-expressing (green, $n = 10$) and non-expressing (gray, $n = 14$) cells to a 5 ms pulse of 473 nm light are shown. The box and whisker plot shows the average response latency for each ChR2-expressing cell ex vivo.

(D) Photoresponse latencies in vivo were calculated by measuring the time from stimulation to 4 SD above the baseline firing rate.

(E) A bimodal distribution of excitatory photoresponse latencies was identified in recorded units ($n = 198$) and divided into Type 1 (green; $n = 19$) and Type 2 units (blue; $n = 34$).

(F) Type 1 units responded to photostimulation with fast excitation (3–8 ms latency). Inset shows the overlaid average traces for spontaneous spiking (black) and light-evoked spiking (blue) from a representative unit.

(G) Type 2 units responded to photostimulation with delayed excitation (80–120 ms latency).

(H) Scatterplot depicting the peak-trough duration of the waveform plotted against the average firing rate for each unit.

(I and J) Normalized histogram showing the distribution of peak-trough durations for Type 1 units (I) and Type 2 units (J).

(K and L) Diagrams illustrating two possible circuit models. (K) Type 1 units project directly from the LH to the VTA, whereas Type 2 units represent a population in the LH that is receiving feedback from the VTA; or (L) Type 2 units represent a population in the LH that is receiving input from collaterals of Type 1 units. Dotted lines indicate the presence of either a monosynaptic or polysynaptic connection.

Scale bar: y axis, 0.2 mV; x axis, 500 μ s. See also Figure S1.

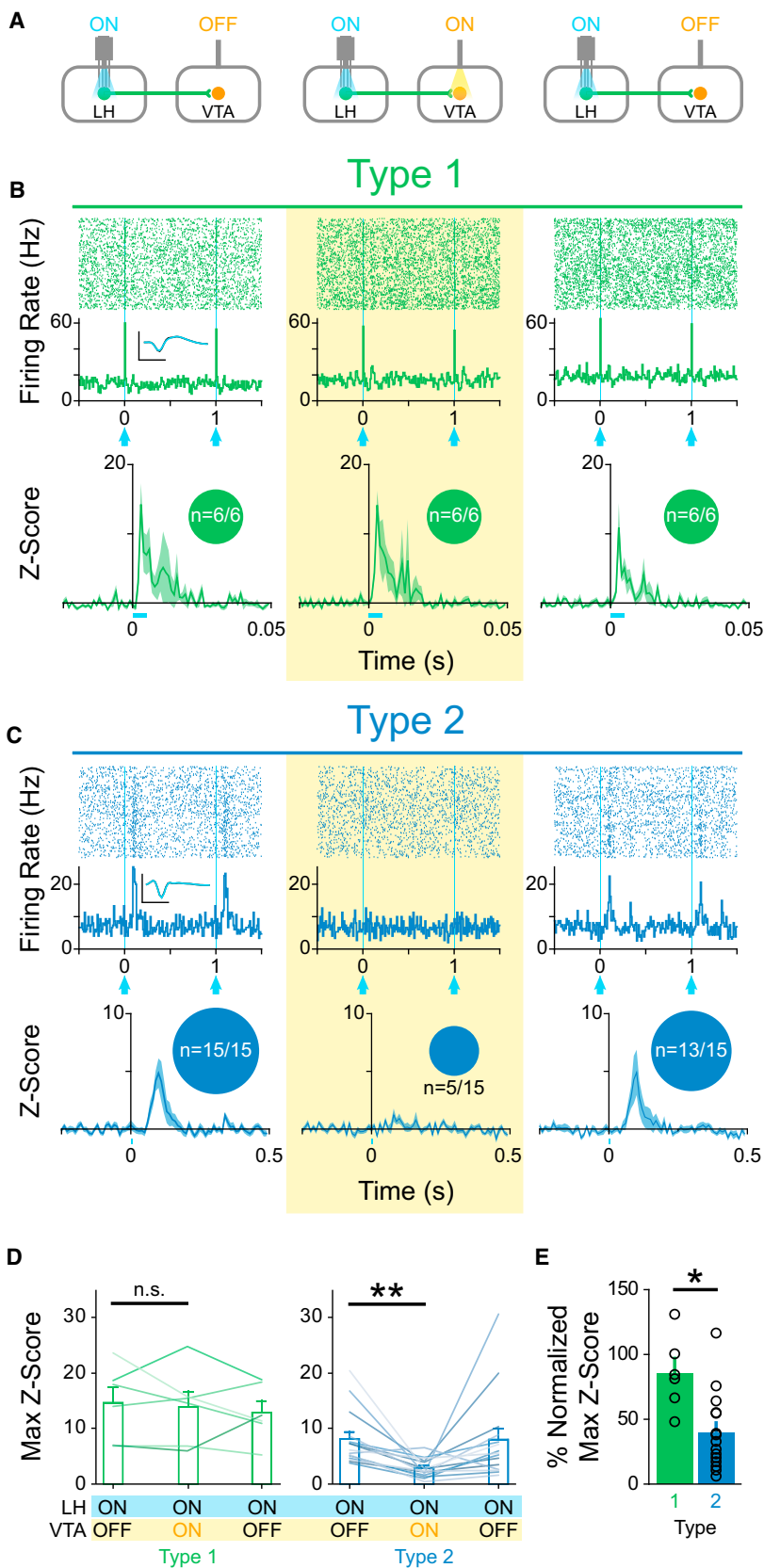


Figure 2. Inhibition of the VTA Selectively Attenuates the Photoresponse of Type 2, but not Type 1, Units

(A) Mice expressing ChR2 in LH-VTA projections received an additional injection of AAV₅-CaMKII α -eNpHR3.0-eYFP into the VTA to allow for transient inhibition of VTA neurons by yellow light. Three epochs of phototagging were conducted (LH photoactivation: ON-ON-ON, VTA photoinhibition: OFF-ON-OFF).

(B) Type 1 ($n = 6/121$ units, $n = 6$ animals) photoresponse properties were unaffected (0%; $n = 0/6$ attenuated or abolished) by VTA inhibition. Inset circles represent the number of units photoresponsive during each epoch. Inset shows the overlaid average traces for spontaneous spiking (black) and light-evoked spiking (blue) from a representative unit.

(C) Type 2 ($n = 15/121$ units, $n = 6$ animals) photoresponse properties were abolished (67%; $n = 10/15$) or attenuated (87%; $n = 13/15$) during NpHR-mediated VTA inhibition.

(D) No significant difference in max Z score was detected between epochs with and without inhibition of the VTA for Type 1 units (two-tailed, paired Student's t test, $p = 0.71$). The max Z score was significantly lower in the ON (LH blue light illumination + VTA photoinhibition) epoch relative to the first OFF epoch (LH blue light illumination only) for Type 2 units (two-tailed, paired Student's t test, $**p = 0.0015$).

(E) There was a significant difference in max Z score (normalized to the OFF epoch) during photoinhibition of the VTA between Type 1 units compared to Type 2 units (two-tailed, unpaired Student's t test, $*p = 0.014$).

Error bars indicate + SEM. Scale bar: y axis, 0.2 mV; x axis, 500 μ s. See also Figure S3.

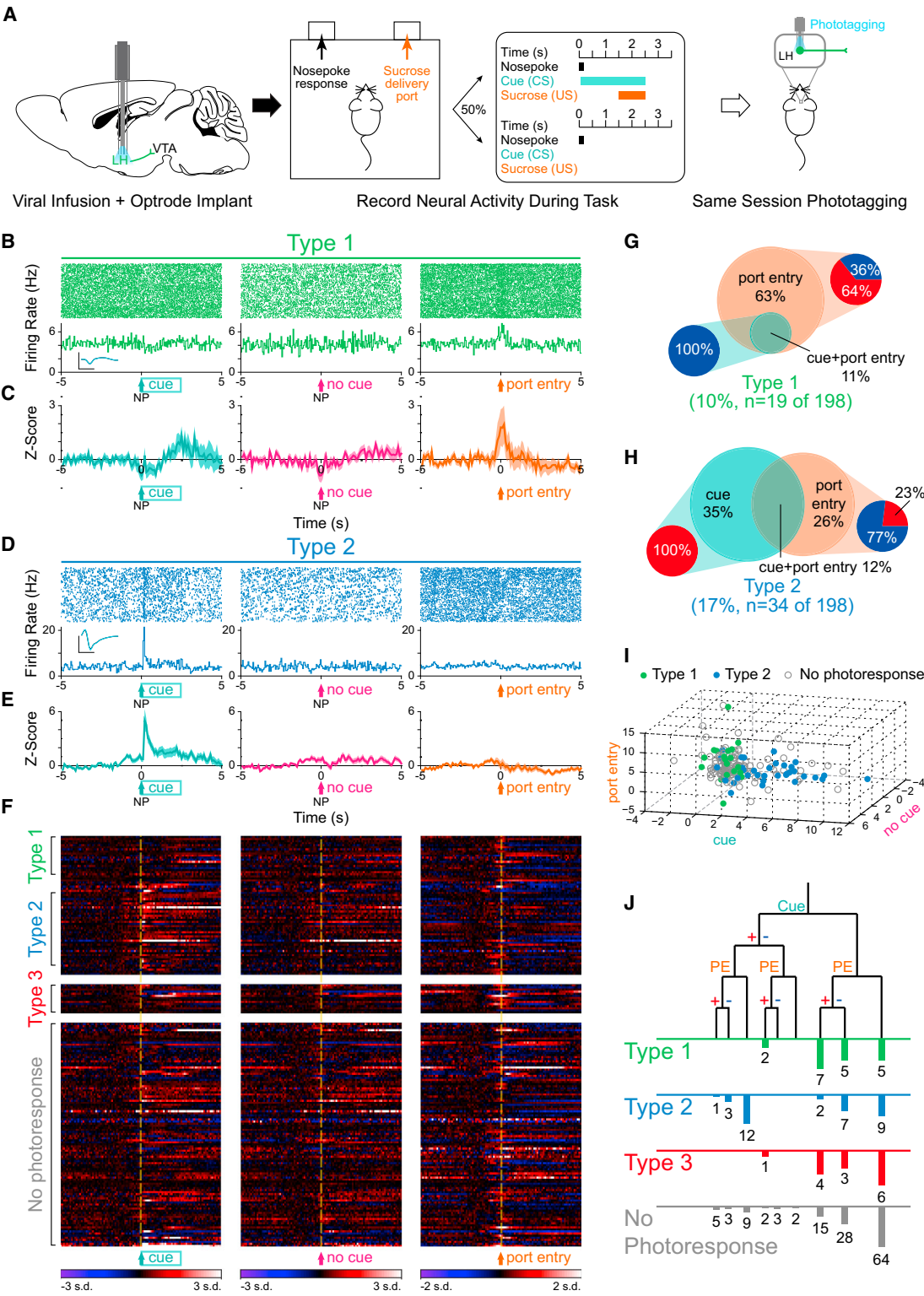


Figure 3. Type 1 Units Predominantly Respond to the Port Entry, whereas Type 2 Units Respond to Both the Conditioned Stimulus and the Port Entry

(A) Mice with optrodes implanted in the LH and expressing ChR2 in LH-VTA projections were trained on a task where 50% of nosepokes (NP) were followed by a cue (conditioned stimulus; CS) that predicts the delivery of sucrose (unconditioned stimulus; US) at the delivery port. In vivo electrophysiological recordings were performed during the behavioral task followed by phototagging in the same recording session to identify units by projection target.

(legend continued on next page)

(74%; $n = 14/19$) were either phasically excited or inhibited by sucrose port entry, with a small number also showing phasic inhibition to the reward-predictive cue (Figures 3B, 3C, and 3G). In contrast, Type 2 units were more heterogeneous, with task-responsive neurons encoding the cue selectively (35%), the sucrose port-entry selectively (26%), or both the cue and port entry (12%; Figures 3D, 3E, and 3H). To illustrate the strength of responses of Type 1 and Type 2 units to task-related events, we plotted each cell on a three-dimensional plot according to Z score (Figure 3I). To show the distribution of phasic changes in firing to multiple task-related events on a qualitative level, we plotted the number of cells of each photoresponse type that fell into a given category (Figure 3J).

Different Components of the LH-VTA Circuit Represent Distinct Aspects of Reward-Related Behavior

Given the well-defined role of the VTA in reward-prediction error (e.g., the phasic reduction of DA neuron firing in response to the unexpected omission of a reward and the phasic excitation in response to unexpected reward delivery) (Schultz et al., 1997), we investigated whether LH neurons would encode the unexpected omission of a sucrose reward. To do this, we recorded the neural activity of photoresponsive neurons during the same cue-reward task in well-trained animals but randomly omitted 30% of sucrose deliveries following the cue (Figure 4A).

The majority of Type 1 units (88%; $n = 15/17$, binomial distribution, $p = 0.001$) were insensitive to reward omission (Figures 4B and 4D), whereas a large subset of Type 2 units (67%; $n = 12/18$) showed a significantly different response to reward-presented and reward-omitted trials (Figures 4C and 4D). We concluded that LH-VTA (Type 1) neurons encoded the action of entering the port, as these port-entry responses were persistent even upon reward omission (Figure 4D), in contrast to Type 2 units ($\chi^2 = 10.9804$, $p = 0.0009$).

To determine whether Type 1 responses to port entry were truly encoding the conditioned response (CR), as opposed to general reward-seeking or exploratory behavior, we recorded in untrained mice that had not yet acquired the task. In task-naïve mice, we delivered sucrose to the port in the absence of a predictive cue (unpredicted reward delivery) and found that Type 1 units did not show phasic responses to port entry (Figures

4E, 4F, and 4I), consistent with the model that Type 1 neurons encode the CR (Figure 4J).

Next, to determine whether Type 2 unit activity is consistent with a reward-prediction error-like response profile, we also recorded these neurons in well-trained animals during unpredicted reward delivery (Figure 4G). We found that a subset of Type 2 units responded to unpredicted sucrose deliveries (50%; Figures 4G–4I). Taken together, subsets of Type 2 units are sensitive to unexpected reward omission (Figures 4C and 4D) and unpredicted reward delivery (Figures 4G–4I), consistent with a reward-prediction error-like response profile.

Photostimulation of the LH-VTA Pathway Promotes Sucrose Seeking in the Face of a Negative Consequence

As we have shown above, Type 1 units represent a neural correlate of CR. Importantly, the increase in firing rate begins prior to CR, ramping up until the CR has been completed (Figures 3B, 3C, and 4B). To determine whether activation of the LH-VTA pathway could promote CR, we wanted to test the ability of LH-VTA activation in driving CR in the face of a negative consequence. In wild-type mice, we expressed ChR2-eYFP or eYFP alone in LH cell bodies and implanted an optic fiber over the VTA (Figures 5A and S4). Conversely, to test the role of the LH-VTA pathway in mediating CR or feeding-related behaviors, we bilaterally expressed NpHR-eYFP or eYFP alone in LH cells and implanted an optic fiber above the VTA (Figures 5A and S4).

We designed a Pavlovian conditioning task in which food-deprived mice had to cross a shock grid to retrieve a sucrose reward (Figure 5B). In the first “baseline” epoch (with the shock grid off), we verified that each mouse had acquired the Pavlovian conditioned approach task. In the second (“Shock”) epoch, the shock grid delivered mild foot shocks every second. Finally, in the third epoch (“Shock+Light”), we continued to deliver foot shocks but also illuminated LH terminals in the VTA with blue light (10 Hz) in mice expressing ChR2 and matched eYFP controls and yellow light (constant) for mice expressing NpHR and their eYFP controls (Figure 5B).

We observed a significantly higher number of port entries per cue during the Shock+Light epoch and a significantly higher difference score (Shock+Light epoch – Shock-only epoch) in ChR2 mice relative to eYFP mice (Figure 5C and Movie S1). In contrast, photoinhibition of the LH-VTA pathway resulted in a significant

(B) Perievent raster histograms for a representative Type 1 unit that responded to port entry, but not to the reward-predictive cue. Inset shows overlaid average traces for spontaneous spiking (black) and light-evoked spiking (blue) from a representative unit.

(C) Population Z score plots showing the average responses of all Type 1 units ($n = 19/198$ units, $n = 12$ animals).

(D) Perievent raster histograms for a representative Type 2 unit that responded to the reward-predictive cue, but not to port entry.

(E) Population Z score plots show the average responses of all Type 2 units ($n = 34/198$ units, $n = 12$ animals).

(F) Heatmap representation of the individual Z scores of all units.

(G) Of all Type 1 units, 63% responded exclusively to the port entry ($n = 12/19$), whereas 11% responded to both the port entry and the reward-predictive cue ($n = 2/19$). Within the Type 1 units that responded to the port entry, 64% ($n = 9/14$) were excited (red) upon port entry, whereas 36% ($n = 5/14$) were inhibited (blue), and within the units that responded to the reward-predictive cue, 100% ($n = 2/2$) were inhibited by the cue.

(H) Of all Type 2 units, 35% ($n = 12/34$) responded exclusively to the reward-predictive cue, 26% ($n = 9/34$) responded exclusively to the port entry, and 12% ($n = 4/34$) responded to both. Within the Type 2 units that responded to the cue, 100% ($n = 16/16$) were excited by the cue, whereas none were inhibited, and within the units that responded to port entry, 77% ($n = 10/13$) were inhibited upon port entry, whereas 23% ($n = 3/13$) were excited.

(I) Graphical representation of Z scores during the experimental windows for cue, no cue, and port entry for Type 1, Type 2, and “no photoresponse” units.

(J) Diagram of recorded units demonstrating whether they responded to the cue or port entry (PE) and whether that response was with excitation (+) or inhibition (–).

Error bars indicate + SEM. Scale bar: y axis, 0.2 mV; x axis, 500 μ s. See also Figure S2.

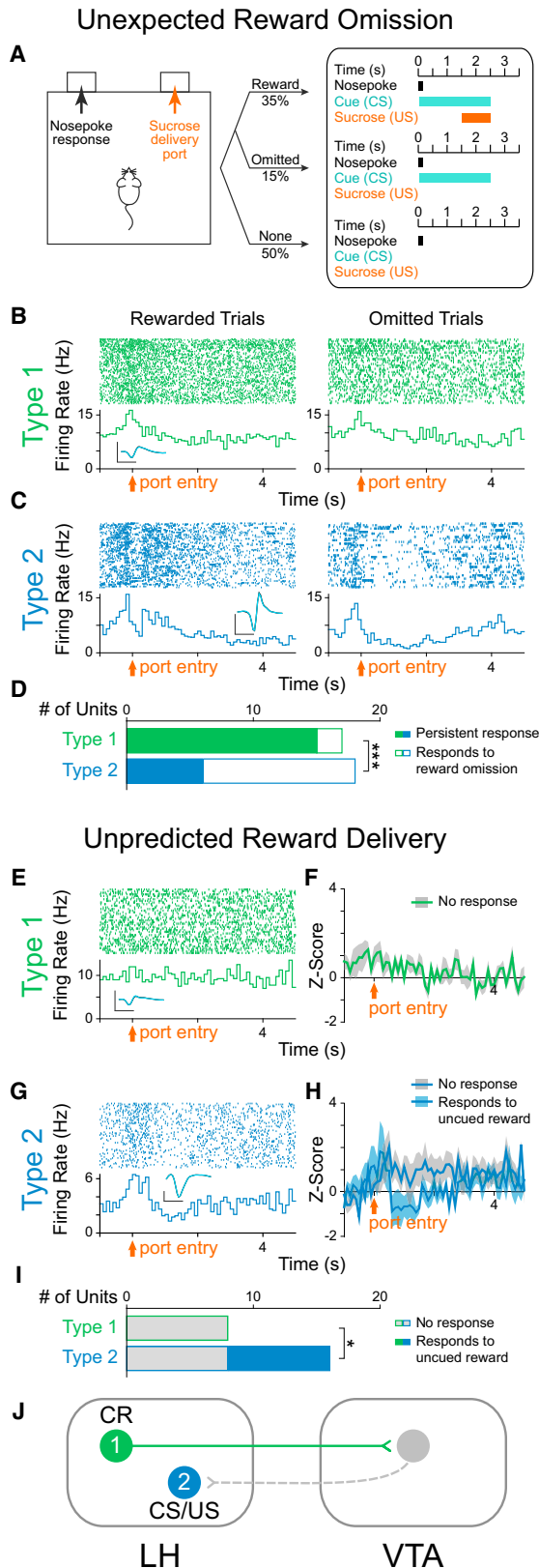


Figure 4. LH-VTA Neurons Encode the CR of Sucrose Seeking

(A) The original partial reinforcement sucrose self-administration task was modified so that in 30% of trials during which the reward-predictive cue was present, the expected sucrose delivery was omitted (15% of all trials). (B) Perievent raster histograms of a Type 1 unit that showed no difference in response to port entry with reward omission. Inset shows overlaid average traces for spontaneous spiking (black) and light-evoked spiking (blue) from a representative unit. (C) Perievent raster histograms of a Type 2 unit that showed a significantly different response to port entry upon omission of the expected reward. (D) Of all Type 1 units recorded ($n = 17/122$ units, $n = 6$ animals), only 12% ($n = 2/17$) showed a significant difference in their responses when the expected reward was omitted. In contrast, of all Type 2 units recorded ($n = 18/122$ units, $n = 6$ animals), 67% ($n = 12/18$) showed a significant difference in their responses when the expected reward was omitted (chi-square = 10.9804, $***p = 0.0009$). (E) Unexpected sucrose delivery occurred in the absence of predictive cues. Perievent raster histogram of a Type 1 unit that did not respond to port entry following unpredicted reward delivery is shown. (F) Population Z score plot showing the average responses of all Type 1 units to the port entry following unpredicted reward delivery. (G) Perievent raster histogram of a Type 2 unit that showed an increase in firing rate to port entry following unpredicted reward delivery. (H) Population Z score plot of Type 2 unit responses to port entry following unpredicted reward delivery, separated into those that showed a significant response and those that showed no significant response. (I) Of all Type 1 units recorded ($n = 8/105$ units, $n = 6$ animals), 0% ($n = 0/8$) showed a significant response to the port entry following unpredicted reward delivery. In contrast, of all Type 2 units recorded ($n = 16/105$ units, $n = 6$ animals), 50% ($n = 8/16$) showed a significant response to the port entry following unpredicted reward delivery (chi-square = 6, $p = 0.0143$). (J) Schematic of the LH-VTA loop and the components of reward processing encoded by Type 1 and 2 cells. CR = conditioned response; CS = conditioned stimulus; US = unconditioned stimulus. Scale bar: y axis, 0.2 mV; x axis, 500 μ s.

reduction in port entries per cue and difference scores in the NpHR mice relative to eYFP mice (Figure 5D and Movie S2). Within-session extinction experiments during which cue presentations were not followed by sucrose deliveries showed similar trends in effect (Figure S4).

Importantly, we wanted to determine whether the changes in sucrose seeking we had obtained were caused by changes in feeding-related behavior or sensitivity to pain. We observed that photoactivation of the LH-VTA projection significantly increased the time spent feeding in well-fed mice in the ChR2 group (Figure 5E). However, photoinhibition of the LH-VTA pathway did not significantly reduce feeding (Figure 5F), even though these animals were food deprived to enhance our ability to detect a reduction relative to the baseline epoch (compare to sated animals in Figure 5E). In neither the ChR2 (Figure 5G) nor NpHR group (Figure 5H) did we observe a difference in latency to tail withdrawal from hot water (Ben-Bassat et al., 1959; Grotto and Sulman, 1967), indicating that manipulating the LH-VTA projection was not altering analgesia.

LH Provides Both Glutamatergic and GABAergic Input onto VTA DA and GABA Neurons

To study the composition of the fast transmission components of LH inputs to the VTA that were eliciting these effects, we performed whole-cell patch-clamp recordings from VTA neurons in an acute slice preparation while optically activating LH

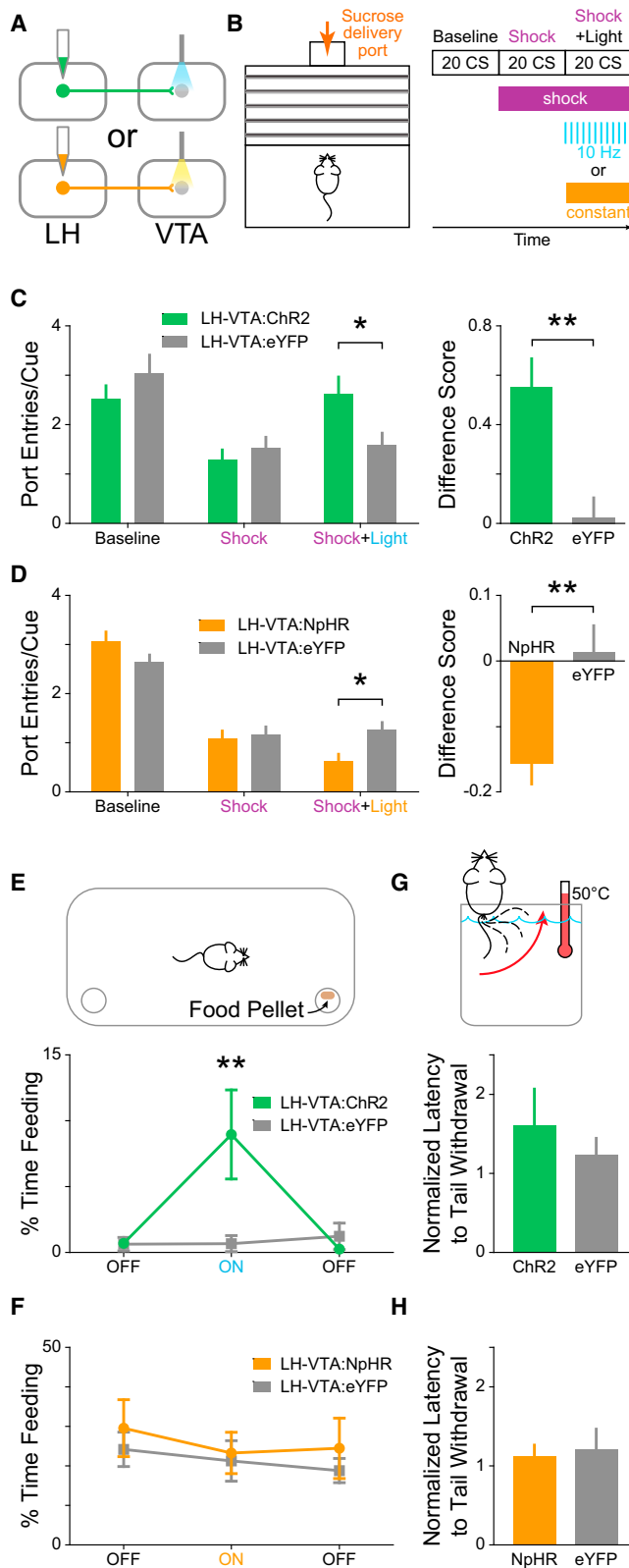


Figure 5. Excitation of LH-VTA Projections Promotes, whereas Inhibition Attenuates, Compulsive Sucrose Seeking

(A) Mice received injections of AAV₅-CaMKII α -ChR2-eYFP ($n = 8$), AAV₅-CaMKII α -eNpHR3.0-eYFP ($n = 14$), or AAV₅-CaMKII α -eYFP ($n = 6$ controls for ChR2, $n = 8$ controls for NpHR) into the LH, and an optic fiber was implanted above the VTA.

(B) Mice were trained on a Pavlovian conditioned approach task wherein a cue predicted sucrose delivery to a port located across a shock grid. On test day, mice were presented with 20 cues during a baseline period without shock, 20 cues when the shock grid was on, and 20 cues during which 10 Hz blue or constant yellow light was delivered while the shock floor remained on.

(C) Mice in the ChR2 group showed a significant increase in the number of port entries per cue during the “Shock+Light” epoch relative to eYFP controls ($n = 8$ ChR2, $n = 6$ eYFP; two-way ANOVA revealed a group \times epoch interaction, $F_{2,24} = 20.47$, $p < 0.0001$; Bonferroni post-hoc analysis, $*p < 0.05$). The difference between the number of port entries per cue during the “Shock+Light” epoch and “Shock” epoch was also significantly different between the ChR2 and eYFP control groups (two-tailed, unpaired Student’s t test, $**p = 0.0090$).

(D) Mice in the NpHR group showed a significant decrease in the number of port entries per cue during the Shock+Light epoch relative to eYFP controls ($n = 13$ NpHR, $n = 8$ eYFP; two-way ANOVA revealed a group \times epoch interaction, $F_{2,38} = 116.63$, $p < 0.0001$; Bonferroni post-hoc analysis, $*p < 0.05$). The difference score was also significantly different between the NpHR-expressing and eYFP control mice (two-tailed, unpaired Student’s t test, $**p = 0.0062$).

(E) Mice were placed into an open chamber with two cups, one containing food and the other without, and behavior in three experimental epochs was recorded (light OFF-ON-OFF). ChR2-expressing mice showed a significant increase in feeding (measured by time spent consuming food) compared with eYFP controls during the epoch paired with blue-light stimulation ($n = 8$ ChR2, $n = 6$ eYFP; two-way ANOVA revealed a group \times epoch interaction, $F_{2,24} = 4.23$, $p = 0.0268$; Bonferroni post-hoc analysis, $**p < 0.01$).

(F) NpHR-expressing mice showed no significant differences from eYFP control mice in time spent feeding in any of the epochs ($n = 9$ NpHR, $n = 7$ eYFP).

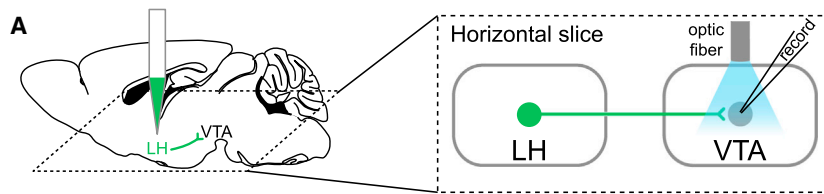
(G and H) To examine the effect of light stimulation on analgesia, mice had their tails placed into a heated water bath, and the latency-to-tail withdrawal was measured during two counterbalanced epochs (light ON-OFF). (G) ChR2-expressing mice showed no significant difference in tail-withdrawal latency (normalized to OFF epoch) during blue-light stimulation compared to eYFP controls ($n = 8$ ChR2, $n = 6$ eYFP), (H) nor did NpHR-expressing mice during yellow-light stimulation ($n = 5$ NpHR, $n = 8$ eYFP).

Error bars indicate \pm SEM. See also Figure S4.

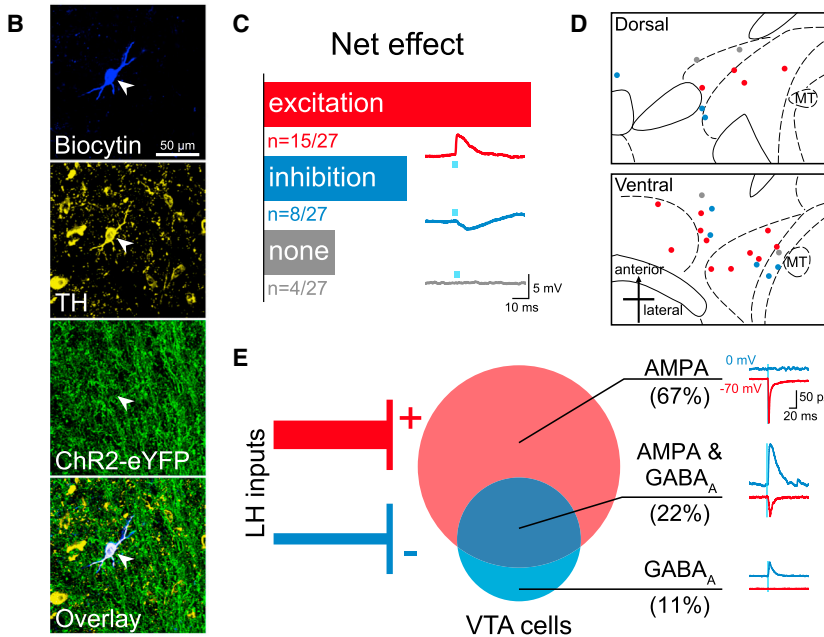
inputs expressing ChR2-eYFP (Figures 6A and S5). Given that there is well-established heterogeneity within the VTA, including $\sim 65\%$ DA neurons, $\sim 30\%$ GABA neurons, and $\sim 5\%$ glutamate neurons (Margolis et al., 2006; Nair-Roberts et al., 2008; Yamaguchi et al., 2007), we filled cells with biocytin while recording to allow for identification of cell type using post-hoc immunohistochemistry for tyrosine hydroxylase (TH; Figure 6B), in addition to recording the hyperpolarization-activated cation current (I_h) and mapping cell location (Figures 6B and S5).

First, we recorded in current-clamp during photostimulation of ChR2-expressing LH inputs and observed that 23 of 27 neurons showed a time-locked response to photostimulation of LH inputs (Figure 6C). The majority of DA neurons sampled in the VTA received a net excitatory input from the LH (56%), whereas another subset showed net inhibition (30%; Figure 6C). The spatial distribution of these DA neurons is mapped onto an atlas for horizontal slices containing the VTA (Figure 6D).

To establish the monosynaptic contribution of LH inputs to VTA DA neurons, we used ChR2-assisted circuit mapping, where voltage-clamp recordings were performed in the presence of tetrodotoxin (TTX) and 4-aminopyridine (4AP; Petreanu



Dopamine neurons



GABA neurons

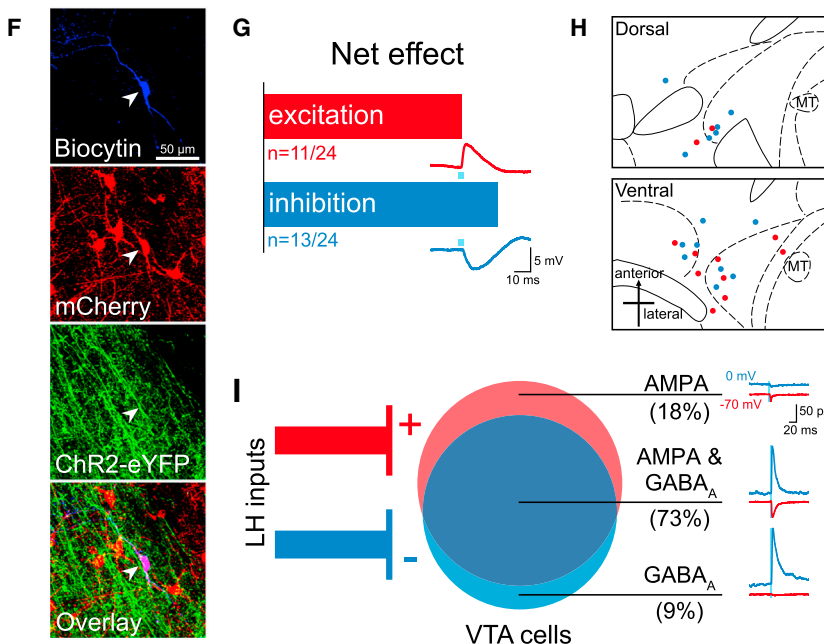


Figure 6. The LH Sends a Mixture of Excitatory and Inhibitory Projections to Both DA and GABA Neurons in the VTA

(A) AAV₅-CaMKII α -ChR2-eYFP was injected into the LH, and at least 6 weeks later, 300 μ m-thick horizontal brain slices were prepared containing the VTA. Whole-cell patch-clamp recordings were made in VTA neurons, and ChR2-expressing LH terminals were activated by illumination with 473 nm light via an optic fiber resting on the brain slice.

(B) Neurons were filled with biocytin during recording, and DA neurons were identified by immunohistochemistry for TH (n = 27).

(C) The net effect of optical stimulation of LH terminals was assessed in current-clamp mode, which revealed that 55% of DA neurons (n = 15/27) showed a net excitatory response, whereas 30% (n = 8/27) responded with net inhibition, and 15% (n = 4/27) showed no response. An example of an excitatory postsynaptic potential (EPSP, red trace), an inhibitory postsynaptic potential (IPSP, blue trace), and a non-responsive cell (gray trace) are shown below each bar.

(D) The distribution of all recorded TH⁺ neurons plotted on horizontal midbrain slices with colors indicating the response to LH terminal photo-stimulation.

(E) VTA DA neurons received only AMPAR-mediated input (67%, n = 6/9), only GABA_AR-mediated input (11%, n = 1/9), or both of these currents (22%, n = 2/9).

(F) VTA GABA neurons were identified by the presence of mCherry (n = 24), achieved by injection of Cre-dependent AAV₅-EF1 α -DIO-mCherry into the VTA of VGAT::Cre mice.

(G) Optical stimulation of LH terminals in current-clamp mode showed that GABA neurons respond with either net excitation (46%, n = 11/24) or net inhibition (54%, n = 13/24) to LH input.

(H) The distribution of each recorded GABA neuron plotted on horizontal midbrain slices with colors indicating the response to LH terminal stimulation.

(I) GABA neurons received a mixture of AMPAR-mediated and GABA_AR-mediated input from the LH (AMPA only: 18%, n = 2/11; AMPA & GABA_A: 73%, n = 8/11; GABA_A: 9%, n = 1/11).

MT = medial terminal nucleus of the accessory optic tract. See also Figures S5 and S6.

et al., 2007). Consistent with our observations from current-clamp recordings, we observed that the majority of recorded VTA DA neurons exclusively received excitatory monosynaptic input from the LH (67%), compared to VTA DA neurons that exclusively received inhibitory monosynaptic input (11%), or both (22%; [Figures 6E and S6](#)).

We identified VTA GABA neurons by injecting a Cre-dependent fluorophore (AAV₅-DIO-mCherry) into the VTA of VGAT::Cre mice and utilized mCherry expression to direct the recording of VTA GABA neurons ($n = 24$; [Figure 6F](#)). Forty-six percent of VTA GABA neurons responded with net excitation, whereas 54% responded with net inhibition, to photostimulation of ChR2-expressing LH inputs ([Figure 6G](#)). The spatial distribution of these cells is shown in [Figure 6H](#). Upon examination of the monosynaptic input from the LH (as described above), we found that 18% of sampled GABA neurons received exclusively excitatory input and 9% received exclusively inhibitory input ([Figure 6I](#)). However, relative to VTA DA neurons, we found that more VTA GABA neurons received both excitatory AMPAR-mediated and inhibitory GABA_AR-mediated monosynaptic input from the LH (73%; chi-square = 5.0505, $p = 0.0246$; [Figures 6I and S6](#)).

Distinct Roles of Glutamatergic and GABAergic Components of the LH-VTA Pathway in Behavior

Given that our ex vivo recordings provided evidence supporting robust input from both GABAergic and glutamatergic LH projections to the VTA, we next probed the role of each component independently. To do this, we used transgenic mouse lines expressing Cre-recombinase in neurons that expressed either vesicular glutamate transporter 2 (VGLUT2) or vesicular GABA transporter (VGAT). We injected AAV₅-DIO-ChR2-eYFP or AAV₅-DIO-eYFP into the LH of VGLUT2::Cre and VGAT::Cre mice and implanted an optic fiber over the VTA ([Figure S7](#)). These animals were then run on each of the behavioral assays shown in [Figure 5](#).

We did not observe any detectable differences in the number of port entries made per cue between mice expressing ChR2 or eYFP in the LH^{glut}-VTA projection ([Figure 7A](#)) or in the LH^{GABA}-VTA projection ([Figure 7B](#)). However upon video analysis, we noticed aberrant gnawing behaviors in the LH^{GABA}-VTA:ChR2 group upon blue-light illumination (see [Movies S3 and S4](#)). In LH^{glut}-VTA mice, although there was a trend toward a reduction in feeding upon photostimulation in the ChR2 group compared to the eYFP group, this was not statistically significant ([Figure 7C](#)). In contrast, we observed a robust increase in the time spent feeding in sated mice upon illumination in the LH^{GABA}-VTA:ChR2 group relative to controls ([Figure 7D](#) and [Movie S3](#)). In neither group of animals was there an effect of light stimulation in the tail-withdrawal assay ([Figures 7E and 7F](#)).

During the feeding task, as we did during the sucrose-seeking task, we again noticed aberrant feeding-related motor sequences that were not directed at food. We filmed a representative mouse in the LH^{GABA}-VTA:ChR2 group in an empty transparent chamber, and upon 20 Hz photostimulation, we observed unusual appetitive motor sequences such as licking and gnawing the floor or empty space ([Movie S4](#)). We quantified these “gnawing” behaviors during the feeding task in the wild-

type LH-VTA ([Figure 7G](#)), LH^{glut}-VTA ([Figure 7H](#)), and LH^{GABA}-VTA ([Figure 7I](#)) groups and showed that LH^{GABA}-VTA:ChR2 mice gnawed more than wild-type or LH^{glut}-VTA:ChR2 mice when photostimulated, as compared to their respective eYFP groups ([Figure 7J](#)). We considered whether the aberrant feeding-related behaviors might be separated from appropriately directed feeding at lower frequencies. However, when we tested the LH^{GABA}-VTA:ChR2 group with 5 Hz and 10 Hz trains of blue light, we observed a proportional relationship between stimulation frequency and both feeding and gnawing ([Figure 7K](#)).

DISCUSSION

Functional Components of the LH-VTA Loop

The LH projection to the VTA has been explored with electrical stimulation collision studies ([Bielajew and Shizgal, 1986](#)) and has long been hypothesized to play a role in reward processing ([Hoebel and Teitelbaum, 1962](#); [Margules and Olds, 1962](#)), yet pinpointing this role has been a challenge. Here, we are providing a detailed dissection of how individual components of the LH-VTA loop process different aspects of a reward-related task.

Through the use of optogenetic-mediated phototagging ([Figure 1](#)), we have identified two separate populations of LH neurons: cells that send projections to the VTA (Type 1) and cells that receive feedback from the VTA (Type 2; [Figure 2](#))—though these populations need not be mutually exclusive, as it is possible that LH neurons could both send and receive inputs to and from the VTA. Interestingly, we found that relatively few photoresponsive neurons fell outside the bimodal distribution encapsulating these two populations ([Figures S2B and 1E](#)). Given this, in combination with the long latency delay in Type 2 photoresponses (~100 ms), we speculate that there may be one dominant pathway contributing to the activity of Type 2 neurons. Additionally, because DA binds G protein-coupled receptors, the kinetics are slower than most glutamatergic synapses ([Girault and Greengard, 2004](#)) and may explain this cluster of 100 ms latency photoresponsive units. It is also possible that the VTA may provide indirect feedback through other distal regions, via excitatory intermediate regions such as the amygdala, or with disinhibition via the nucleus accumbens (NAc) or bed nucleus of the stria terminalis (BNST).

Interestingly, whereas photostimulation of Type 1 units evokes excitatory responses in Type 2 units, Type 1 and 2 units show distinct behavioral encoding properties. For example, the numbers of Type 1 and Type 2 units that selectively encode the reward-predictive cue are significantly different ($n = 0/19$ Type 1 versus $n = 12/34$ Type 2, chi-square = 8.67, $p = 0.003$). This paradoxical response pattern could be due to computational processes at an intermediate circuit element, such as the VTA, that may be playing an active role during the behavioral task but inactive during phototagging. Additionally, the behavioral state of the animal could influence how these data are processed.

Decoding Circuit Components in Reward Processing

Our reward omission experiments allowed us to distinguish between LH neural encoding of the CR and the consumption of

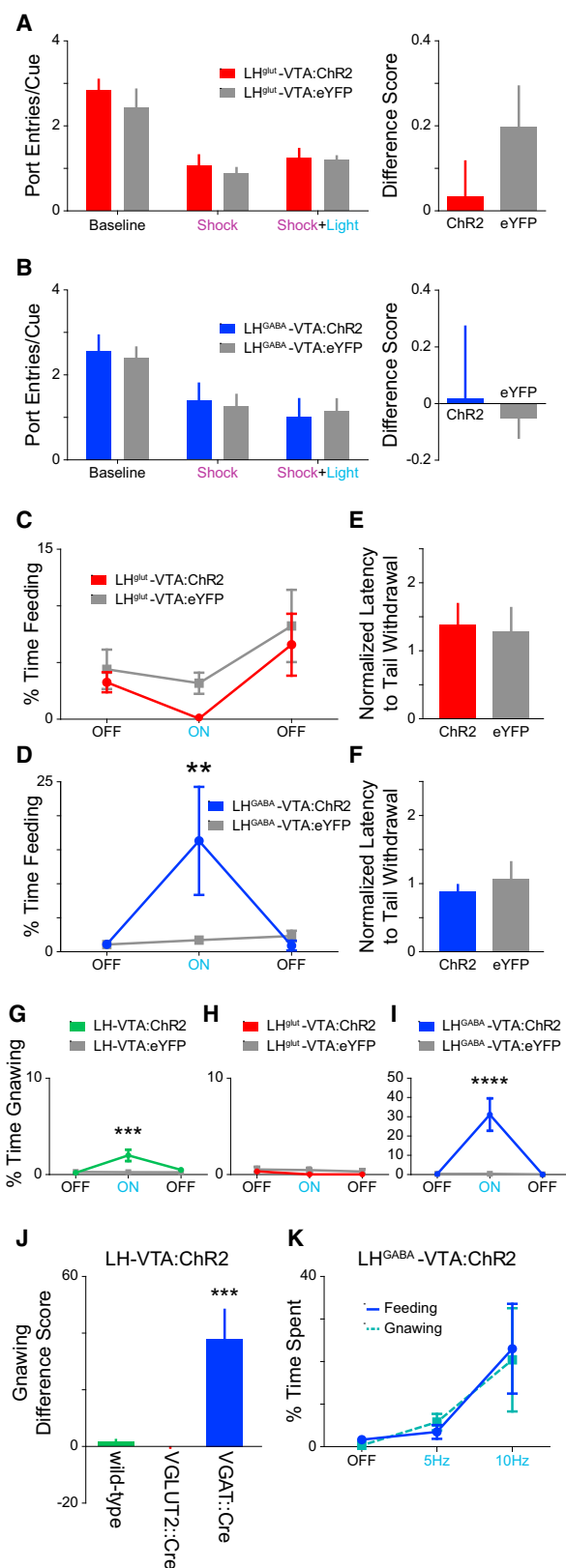


Figure 7. Photoactivation of the GABAergic, but Not the Glutamatergic, Component of the LH-VTA Projection Increased Feeding Behaviors

(A and B) In order to selectively activate glutamatergic or GABAergic LH-VTA projections, VGLUT2::Cre and VGAT::Cre mice received an injection of AAV₅-DIO-ChR2-eYFP or AAV₅-DIO-eYFP into the LH and had an optic fiber implanted over the VTA. In the sucrose-seeking task, there were no significant differences in the numbers of port entries per cue in any epoch for LH^{glut}-VTA:ChR2 mice (n = 7) compared to LH^{glut}-VTA:eYFP control mice (n = 6) (A) nor in those of LH^{GABA}-VTA:ChR2 mice (n = 6) compared to LH^{GABA}-VTA:eYFP control mice (n = 8) (B).

(C) There was no significant difference between LH^{glut}-VTA:ChR2 mice and eYFP controls in feeding behavior.

(D) However, LH^{GABA}-VTA:ChR2 mice showed a significant increase in time spent feeding during light stimulation compared to LH^{GABA}-VTA:eYFP controls (two-way ANOVA revealed a group x epoch interaction, $F_{2,24} = 4.78$, $p = 0.0178$; Bonferroni post-hoc analysis, $^{**}p < 0.01$).

(E and F) Neither LH^{glut}-VTA:ChR2 mice (E) nor LH^{GABA}-VTA:ChR2 mice (F) showed a difference in tail withdrawal latency compared to their respective controls.

(G) LH-VTA:ChR2 mice showed a significant increase in time spent gnawing during the light ON epoch compared to eYFP controls (two-way ANOVA revealed a group x epoch interaction, $F_{2,24} = 4.78$, $p = 0.0179$; Bonferroni post-hoc analysis, $^{***}p < 0.001$).

(H) There was no significant difference between LH^{glut}-VTA:ChR2 and LH^{glut}-VTA:eYFP controls in gnawing behavior.

(I) However, LH^{GABA}-VTA:ChR2 animals also showed a significant increase in time spent gnawing during the light ON epoch compared to LH^{GABA}-VTA:eYFP controls (two-way ANOVA revealed a group x epoch interaction, $F_{2,24} = 18.91$, $p < 0.0001$; Bonferroni post-hoc analysis, $^{****}p < 0.0001$).

(J) The difference score for gnawing behavior between the ON and OFF epochs was significantly greater in LH^{GABA}-VTA:ChR2 animals in comparison with either wild-type LH-VTA:ChR2 or LH^{glut}-VTA:ChR2 animals (one-way ANOVA, $F_{2,18} = 16.76$, $p < 0.0001$; Bonferroni post-hoc analysis, $^{***}p < 0.001$).

(K) Frequency-response curve showing the effect of different blue-light stimulation frequencies (OFF, 5 Hz, 10 Hz) on behavior in LH^{GABA}-VTA:ChR2 animals.

Error bars indicate \pm SEM. See also Figure S7.

the unconditioned stimulus (US). In these experiments, a subset of Type 2 units responded to the reward-predictive cue (CS) and the US and also showed a decrease in firing rate when expected rewards were omitted. Furthermore, a subset of Type 2 units also show phasic excitation upon unexpected reward delivery (Figures 4G and 4H). These data are reminiscent of the way DA neurons in the VTA encode reward-prediction error (Cohen et al., 2012; Schultz et al., 1997). We speculate that VTA neurons may transmit reward-prediction error signals to a subset of LH neurons, which are well-positioned to integrate these signals for the determination of an appropriate behavioral output. Specifically, the LH is robustly interconnected with a multitude of other brain areas (Berthoud and Münzberg, 2011) and has been causally linked to homeostatic states such as sleep/arousal and hunger/satiety (Carter et al., 2009; Jennings et al., 2013).

A Causal Role for the LH-VTA Pathway in Compulsive Sucrose Seeking?

Compulsive reward-seeking behavior has primarily been discussed in the context of drug addiction, wherein a classic paradigm for compulsive drug seeking has been to examine the degree to which drug-seeking behavior persists in the face of a negative consequence, such as a foot shock (Belin et al., 2008; Pelloux et al., 2007; Vanderschuren and Everitt, 2004). We

adapted this task for sucrose seeking to allow us to investigate whether activation of the LH-VTA pathway was sufficient to promote compulsive sucrose seeking. Given that a distinct difference between drug and natural reward is that drug rewards are not necessary for survival, there is controversy as to what behaviors would constitute compulsive sucrose- or food-seeking behavior. An alternative interpretation of our data is that activation of the LH-VTA pathway simply increases motivational drive or the urge to seek appetitive reinforcers. As the rates of obesity have increased in recent decades (Mietus-Snyder and Lustig, 2008), compulsive overeating and sugar addiction are prevalent conditions that are a major threat to human health (Avena, 2007). The feeding behavior in sated (fully fed) mice after activation of the LH-VTA pathway is reminiscent of eating behaviors seen in humans diagnosed with compulsive overeating disorder (or binge-eating disorder) (DSM-V).

It has been proposed that repeated actions lead to the formation of habits, which themselves lead to the compulsive reward seeking that characterizes addiction (Everitt and Robbins, 2005). Our finding that LH-VTA neurons only encode port entry after conditioning suggests that this pathway is selectively encoding a conditioned response, not just a motivated action. This is consistent with our observations that optically activating this projection can promote compulsive reward seeking in the face of a negative consequence (Figure 5C), as well as in the absence of need (as seen in sated mice, Figure 5E). This interpretation is further substantiated by our finding that photoinhibition of the LH-VTA pathway selectively reduces compulsive sucrose seeking (Figure 5D) but does not reduce feeding in food-restricted mice (Figure 5F). One of the greatest challenges in treating compulsive overeating or binge-eating disorders is the risk of impairing feeding behaviors in general. From a translational perspective, we may have identified a specific neural circuit as a potential target for the development of therapeutic interventions for compulsive overeating or sugar addiction without sacrificing natural feeding behaviors.

Composition of LH Input to the VTA

We show that in addition to a glutamatergic LH-VTA component (Kempadoo et al., 2013), there is also a significant GABAergic component in the projection (Leininger et al., 2009), and that LH neurons synapse directly onto both DA and GABA neurons in the VTA (Figure 6). However, there is a difference in the balance of the excitatory/inhibitory input onto VTA DA and GABA neurons.

While we used immunohistochemical processing to verify the identity of VTA neurons, we also measured I_h , a hyperpolarization-activated inwardly rectifying non-specific cation current (Lacey et al., 1989; Ungless and Grace, 2012). The presence of this current has been widely used in electrophysiological studies to identify DA neurons, but it has been shown to be present only in subpopulations of DA neurons, delineated by projection target (Lammel et al., 2011). Although it has previously been proposed in a review by Fields and colleagues that “LH neurons synapse onto VTA projections to the PFC, but not those projecting to the NAc” (Fields et al., 2007), our data suggest that this controversy be reopened for further investigation. Even though we did

observe a subset of DA neurons that received net excitation from the LH and possessed a very small I_h (consistent with mPFC- or NAc medial shell-projecting DA neurons), we also observed a subset of DA neurons that received net excitatory input and showed a large I_h (consistent with characteristics of DA neurons projecting to the lateral shell of the NAc; Figure S5; Lammel et al., 2011). Conversely, VTA DA neurons that received a net inhibitory input showed a very small I_h or lacked this current, which is consistent with the notion that the LH sends predominantly inhibitory input onto VTA DA neurons projecting to the mPFC or the medial shell of the NAc. We also show that LH inputs can be observed in both medial and lateral VTA, suggesting that the LH provides inputs onto VTA neurons with diverse projection targets, as it is known that VTA projection target corresponds somewhat to spatial location along a medial-lateral axis (Lammel et al., 2008).

Excitation/Inhibition Balance in the LH-VTA Pathway

The role of the LH-VTA pathway in promoting reward has previously been ascribed to glutamatergic transmission in the VTA (Kempadoo et al., 2013), as the CaMKII α promoter is often thought to be selective for excitatory projection neurons. However, our data clearly show that expressing ChR2 under the control of the CaMKII α promoter also targets GABAergic projection neurons in the LH (Figure 6).

The behavior elicited by photostimulation of the LH^{GABA}-VTA pathway was frenzied, mis-directed, and maladaptive (Movie S4). One interpretation is that activation of the LH^{GABA}-VTA pathway sends a signal to the mouse that causes the recognition of an appetitive reinforcer. An alternative interpretation is that the LH^{GABA}-VTA pathway might drive incentive salience or an intense “wanting,” consistent with a signal underlying conditioned approach, but at a non-physiological level that produces this aberrant feeding-related behavior (Berridge and Robinson, 2003). Consistent with this, it is possible that activation of the LH^{GABA}-VTA projection actually produces intense sensations of craving, or urges to feed. However, our experiments show that activation of LH^{GABA}-VTA does not produce an increase in compulsive sucrose seeking, but this is likely due to the excessive gnawing and aberrant appetitive behaviors focused on non-food objects in the testing chamber. Although it is difficult to determine the experience of the mouse during this manipulation, it is clear that appropriately directed feeding-related behaviors require the coordinated activation of both the GABAergic and glutamatergic components of the LH-VTA pathway.

Conclusion

Optogenetic and pharmacogenetic manipulations are powerful tools for establishing causal relationships, yet they do not reveal the endogenous, physiological properties of neural circuit elements. Our study unifies information about the synaptic connectivity, the naturally occurring endogenous function, and the causal role of the LH-VTA pathway, providing a new level of insight toward how information is integrated in this circuit. These results highlight the importance of examining the functional role of neurons by connectivity, in addition to genetic markers. LH-VTA neurons selectively encoded the action of reward seeking

but did not encode environmental stimuli, whereas rewarding stimuli and reward-predictive cues were encoded by a discrete population of LH neurons downstream of the VTA. Furthermore, we have identified a specific projection that is causally linked to compulsive sucrose-seeking and feeding behavior. The heterogeneity in the LH-VTA projection is necessary for providing an adaptive balance between driving motivation and regulating appropriately directed appetitive behaviors. These findings provide insights relevant to pathological conditions such as compulsive overeating disorder, sugar addiction, and obesity.

EXPERIMENTAL PROCEDURES

Phototagging VTA-Projecting LH Neurons

To limit expression of ChR2 to only LH neurons projecting to the VTA, AAV₅-DIO-ChR2-eYFP was injected into the LH and HSV-EF1 α -IRES-Cre-mCherry into the VTA. In NpHR inhibition experiments, AAV₅-CaMKII α -eNpHR3.0-eYFP was injected into the VTA as well. An optrode was implanted in the LH and an optic fiber over the VTA.

Partial Reinforcement Sucrose Retrieval Task

For in vivo recording, animals were trained on a partial reinforcement sucrose retrieval task, where 50% of nosepokes were followed by a cue predicting the delivery of sucrose at the port entry. Adjustments were made to this task to examine the effects on reward omission by omitting sucrose deliveries from a subset of cues and to examine the effects on unexpected reward by the delivery of sucrose without the existence of the cue.

Sucrose Seeking in the Face of a Negative Consequence

To study the effect on conditioned responding by stimulation of LH-VTA projections, we developed a task wherein an animal must cross a shock floor to obtain a sucrose reward. Wild-type animals with ChR2, NpHR, or eYFP injected either unilaterally (AAV₅-CaMKII α -ChR2-eYFP) or bilaterally (AAV₅-CaMKII α -eNpHR3.0-eYFP) in the LH with an optic fiber placed over VTA or VGLUT2::Cre and VGAT::Cre animals with AAV₅-DIO-ChR2-eYFP injection in the LH and optic fiber over the VTA were tested. Because LH-VTA:ChR2 mice showed an increase in sucrose seeking in the face of a negative consequence, these animals were sated before evaluating the effects of photostimulation on feeding on normal chow. In contrast, LH-VTA:NpHR mice showed a decrease in sucrose seeking in the face of a negative consequence and were therefore mildly food restricted before testing the effects of photostimulation on feeding on normal chow.

Ex Vivo Characterization of LH-VTA

Whole-cell patch-clamp recordings were used to study the input of LH neurons onto DA and GABA VTA neurons. DA neurons were identified by filling cells with biocytin and post-hoc immunostaining for TH. GABA cells were identified during recordings by fluorescence due to AAV₅-DIO-mCherry injection into the VTA of VGAT::Cre animals.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Discussion, Extended Experimental Procedures, seven figures, and four movies and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.01.003>.

AUTHOR CONTRIBUTIONS

E.H.N. and G.A.M. performed electrophysiological recordings and analyses for in vivo and ex vivo experiments, respectively. S.A.A., E.H.N., K.N.P., and C.A.L. performed behavioral experiments. R.W., K.N.P., C.A.L., and E.H.N. performed histological verification. R.N. provided HSV virus. K.M.T. and C.P.W. supervised experiments and trained experimentalists. E.H.N., G.A.M., S.A.A., and K.M.T. designed experiments. E.H.N. and K.M.T. wrote

the manuscript; all authors contributed to the editing and revision of manuscript.

ACKNOWLEDGMENTS

We thank N. Golan, R. Thomas, M. Anahtar, G. Glober, and A. Beyeler for their assistance with immunohistochemistry. We would also like to thank C. Seo, and S. Kim for their contributions throughout the project and M. Wilson and P. Shizgal for helpful discussion. K.M.T. is a New York Stem Cell Foundation - Robertson Investigator and acknowledges funding from the JPB Foundation, PIIF, PNDRF, Whitehall Foundation, Klingenstein Foundation, NARSAD Young Investigator Award, Alfred P. Sloan Foundation, Whitehead Career Development Chair, NIH R01-MH102441-01 (NIMH), and NIH Director's New Investigator Award DP2-DK-102256-01 (NIDDK). E.H.N. was supported by the NSF Graduate Research Fellowship, the Integrative Neuronal Systems Fellowship, and the Training Program in the Neurobiology of Learning and Memory. G.A.M. was supported by the Simons Center for the Social Brain Postdoctoral Fellowship. S.A.A. was supported by the Jeffrey and Nancy Halis Fellowship as well as the Henry E. Singleton Fund. C.A.L. was supported by the Integrative Neuronal Systems Fellowship and the James R. Killian Fellowship. R.W. was supported by the Netherlands Organisation for Scientific Research (NWO) RUBICON fellowship program.

Received: July 22, 2014

Revised: November 2, 2014

Accepted: December 23, 2014

Published: January 29, 2015

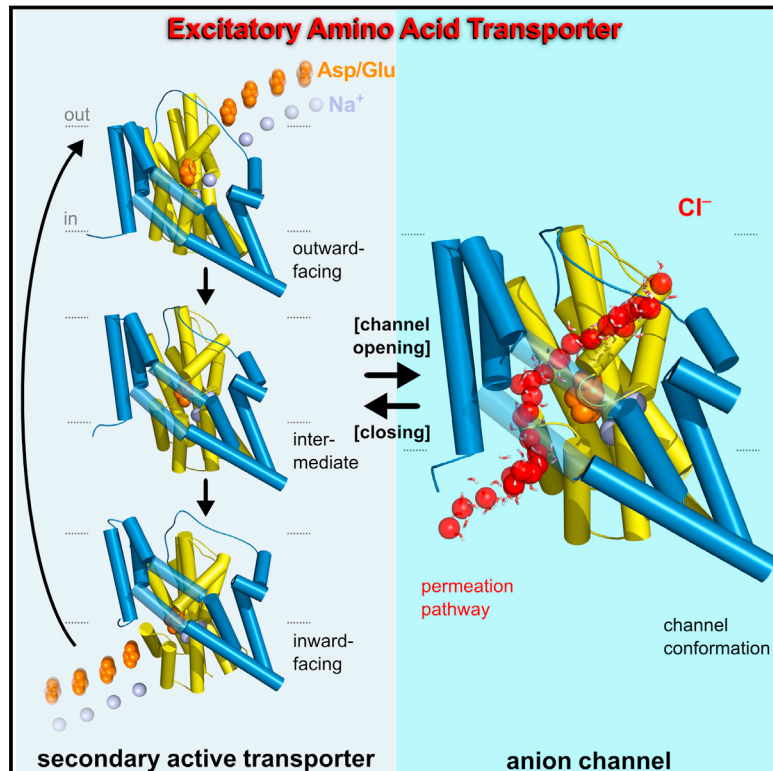
REFERENCES

- Avena, N.M.; Experimental and Clinical Psychopharmacology (2007). Examining the addictive-like properties of binge eating using an animal model of sugar dependence. *Exp. Clin. Psychopharmacol.* 15, 481–491.
- Barone, F.C., Wayner, M.J., Scharoun, S.L., Guevara-Aguilar, R., and Aguilar-Baturoni, H.U. (1981). Afferent connections to the lateral hypothalamus: a horseradish peroxidase study in the rat. *Brain Res. Bull.* 7, 75–88.
- Beckstead, R.M., Domesick, V.B., and Nauta, W.J. (1979). Efferent connections of the substantia nigra and ventral tegmental area in the rat. *Brain Res.* 175, 191–217.
- Belin, D., Mar, A.C., Dalley, J.W., Robbins, T.W., and Everitt, B.J. (2008). High impulsivity predicts the switch to compulsive cocaine-taking. *Science* 320, 1352–1355.
- Ben-Bassat, J., Peretz, E., and Sulman, F.G. (1959). Analgesimetry and ranking of analgesic drugs by the receptacle method. *Arch. Int. Pharmacodyn. Ther.* 122, 434–447.
- Berridge, K.C., and Robinson, T.E. (2003). Parsing reward. *Trends Neurosci.* 26, 507–513.
- Berthoud, H.-R., and Münzberg, H. (2011). The lateral hypothalamus as integrator of metabolic and environmental needs: from electrical self-stimulation to opto-genetics. *Physiol. Behav.* 104, 29–39.
- Bielajew, C., and Shizgal, P. (1986). Evidence implicating descending fibers in self-stimulation of the medial forebrain bundle. *J. Neurosci.* 6, 919–929.
- Burton, M.J., Rolls, E.T., and Mora, F. (1976). Effects of hunger on the responses of neurons in the lateral hypothalamus to the sight and taste of food. *Exp. Neurol.* 51, 668–677.
- Carter, M.E., Adamantidis, A., Ohtsu, H., Deisseroth, K., and de Lecea, L. (2009). Sleep homeostasis modulates hypocretin-mediated sleep-to-wake transitions. *J. Neurosci.* 29, 10939–10949.
- Cohen, J.Y., Haesler, S., Vong, L., Lowell, B.B., and Uchida, N. (2012). Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* 482, 85–88.
- Everitt, B.J., and Robbins, T.W. (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nat. Neurosci.* 8, 1481–1489.

- Fields, H.L., Hjelmstad, G.O., Margolis, E.B., and Nicola, S.M. (2007). Ventral tegmental area neurons in learned appetitive behavior and positive reinforcement. *Annu. Rev. Neurosci.* 30, 289–316.
- Girault, J.A., and Greengard, P. (2004). The neurobiology of dopamine signaling. *Arch. Neurol.* 61, 641–644.
- Gratton, A., and Wise, R.A. (1988). Comparisons of refractory periods for medial forebrain bundle fibers subserving stimulation-induced feeding and brain stimulation reward: a psychophysical study. *Brain Res.* 438, 256–263.
- Grotto, M., and Sulman, F.G. (1967). Modified receptacle method for animal analgesimetry. *Arch. Int. Pharmacodyn. Ther.* 165, 152–159.
- Hoebel, B.G., and Teitelbaum, P. (1962). Hypothalamic control of feeding and self-stimulation. *Science* 135, 375–377.
- Jennings, J.H., Rizzi, G., Stamatakis, A.M., Ung, R.L., and Stuber, G.D. (2013). The inhibitory circuit architecture of the lateral hypothalamus orchestrates feeding. *Science* 341, 1517–1521.
- Kempadoo, K.A., Tourino, C., Cho, S.L., Magnani, F., Leininger, G.-M., Stuber, G.D., Zhang, F., Myers, M.G., Deisseroth, K., de Lecea, L., and Bonci, A. (2013). Hypothalamic neurotensin projections promote reward by enhancing glutamate transmission in the VTA. *J. Neurosci.* 33, 7618–7626.
- Lacey, M.G., Mercuri, N.B., and North, R.A. (1989). Two cell types in rat substantia nigra zona compacta distinguished by membrane properties and the actions of dopamine and opioids. *J. Neurosci.* 9, 1233–1241.
- Lammel, S., Hetzel, A., Häckel, O., Jones, I., Liss, B., and Roeper, J. (2008). Unique properties of mesoprefrontal neurons within a dual mesocorticolimbic dopamine system. *Neuron* 57, 760–773.
- Lammel, S., Ion, D.I., Roeper, J., and Malenka, R.C. (2011). Projection-specific modulation of dopamine neuron synapses by aversive and rewarding stimuli. *Neuron* 70, 855–862.
- Leininger, G.M., Jo, Y.-H., Leshan, R.L., Louis, G.W., Yang, H., Barrera, J.G., Wilson, H., Opland, D.M., Faouzi, M.A., Gong, Y., et al. (2009). Leptin acts via leptin receptor-expressing lateral hypothalamic neurons to modulate the mesolimbic dopamine system and suppress feeding. *Cell Metab.* 10, 89–98.
- Margolis, E.B., Lock, H., Hjelmstad, G.O., and Fields, H.L. (2006). The ventral tegmental area revisited: is there an electrophysiological marker for dopaminergic neurons? *J. Physiol.* 577, 907–924.
- Margules, D.L., and Olds, J. (1962). Identical “feeding” and “rewarding” systems in the lateral hypothalamus of rats. *Science* 135, 374–375.
- Mietus-Snyder, M.L., and Lustig, R.H. (2008). Childhood obesity: adrift in the “limbic triangle”. *Annu. Rev. Med.* 59, 147–162.
- Nair-Roberts, R.G., Chatelain-Badie, S.D., Benson, E., White-Cooper, H., Bolam, J.P., and Ungless, M.A. (2008). Stereological estimates of dopaminergic, GABAergic and glutamatergic neurons in the ventral tegmental area, substantia nigra and retrorubral field in the rat. *Neuroscience* 152, 1024–1031.
- Nakamura, K., Ono, T., and Tamura, R. (1987). Central sites involved in lateral hypothalamus conditioned neural responses to acoustic cues in the rat. *J. Neurophysiol.* 58, 1123–1148.
- Norgren, R. (1970). Gustatory responses in the hypothalamus. *Brain Res.* 21, 63–77.
- Olds, J., and Milner, P. (1954). Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *J. Comp. Physiol. Psychol.* 47, 419–427.
- Paxinos, G., and Franklin, K.B.J. (2001). *The Mouse Brain in Stereotaxic Coordinates*, Second Edition (New York: Academic Press).
- Pelloux, Y., Everitt, B.J., and Dickinson, A. (2007). Compulsive drug seeking by rats under punishment: effects of drug taking history. *Psychopharmacology (Berl.)* 194, 127–137.
- Peteanu, L., Huber, D., Sobczyk, A., and Svoboda, K. (2007). Channelrhodopsin-2-assisted circuit mapping of long-range callosal projections. *Nat. Neurosci.* 10, 663–668.
- Peteanu, L., Mao, T., Sternson, S.M., and Svoboda, K. (2009). The subcellular organization of neocortical excitatory connections. *Nature* 457, 1142–1145.
- Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. *Science* 275, 1593–1599.
- Schwartzbaum, J.S. (1988). Electrophysiology of taste, feeding and reward in lateral hypothalamus of rabbit. *Physiol. Behav.* 44, 507–526.
- Simon, H., Le Moal, M., and Calas, A. (1979). Efferents and afferents of the ventral tegmental-A10 region studied after local injection of [³H]leucine and horseradish peroxidase. *Brain Res.* 178, 17–40.
- Singh, J., Desiraju, T., and Raju, T.R. (1996). Comparison of intracranial self-stimulation evoked from lateral hypothalamus and ventral tegmentum: analysis based on stimulation parameters and behavioural response characteristics. *Brain Res. Bull.* 41, 399–408.
- Tabuchi, E., Yokawa, T., Mallick, H., Inubushi, T., Kondoh, T., Ono, T., and Torii, K. (2002). Spatio-temporal dynamics of brain activated regions during drinking behavior in rats. *Brain Res.* 951, 270–279.
- Tye, K.M., Stuber, G.D., de Ridder, B., Bonci, A., and Janak, P.H. (2008). Rapid strengthening of thalamo-amygdala synapses mediates cue-reward learning. *Nature* 453, 1253–1257.
- Ungless, M.A., and Grace, A.A. (2012). Are you or aren't you? Challenges associated with physiologically identifying dopamine neurons. *Trends Neurosci.* 35, 422–430.
- Vanderschuren, L.J.M.J., and Everitt, B.J. (2004). Drug seeking becomes compulsive after prolonged cocaine self-administration. *Science* 305, 1017–1019.
- Watabe-Uchida, M., Zhu, L., Ogawa, S.K., Vamanrao, A., and Uchida, N. (2012). Whole-brain mapping of direct inputs to midbrain dopamine neurons. *Neuron* 74, 858–873.
- Wise, R.A. (2004). Dopamine, learning and motivation. *Nat. Rev. Neurosci.* 5, 483–494.
- Yamaguchi, T., Sheen, W., and Morales, M. (2007). Glutamatergic neurons are present in the rat ventral tegmental area. *Eur. J. Neurosci.* 25, 106–118.
- Yamamoto, T., Matsuo, R., Kiyomitsu, Y., and Kitamura, R. (1989). Response properties of lateral hypothalamic neurons during ingestive behavior with special reference to licking of various taste solutions. *Brain Res.* 481, 286–297.
- Zhang, S.-J., Ye, J., Miao, C., Tsao, A., Cerniauskas, I., Ledergerber, D., Moser, M.-B., and Moser, E.I. (2013). Optogenetic dissection of entorhinal-hippocampal functional connectivity. *Science* 340, 1232627.

Mechanisms of Anion Conduction by Coupled Glutamate Transporters

Graphical Abstract



Authors

Jan-Philipp Machtens, Daniel Kortzak, ..., Rodolfo Briones, Christoph Fahlke

Correspondence

jan-philipp.machtens@gmx.de (J.-P.M.),
c.fahlke@fz-juelich.de (Ch.F.)

In Brief

Excitatory amino acid transporters operate both as transporters and as anion-selective ion channels at synapses. A combination of simulations and experiments with prokaryotic and mammalian glutamate transporter homologs defines the anion conduction pathway and elucidates how a class of secondary active transporters can function as perfectly selective, gated anion channels.

Highlights

- Molecular dynamics simulations define anion-conducting transporter conformations
- Anion permeation occurs along a well-defined, conserved permeation pathway
- Transport intermediates open the channel via steric and hydrophobic gating
- Anion selectivity is achieved via a single, structurally conserved arginine



Mechanisms of Anion Conduction by Coupled Glutamate Transporters

Jan-Philipp Machtens,^{1,2,3,*} Daniel Kortzak,¹ Christine Lansche,¹ Ariane Leinenweber,² Petra Kilian,² Birgit Begemann,² Ulrich Zachariae,⁴ David Ewers,² Bert L. de Groot,³ Rodolfo Briones,³ and Christoph Fahlke^{1,*}

¹Institute of Complex Systems, Zelluläre Biophysik (ICS-4), Forschungszentrum Jülich, 52428 Jülich, Germany

²Institut für Neurophysiologie, Medizinische Hochschule Hannover, 30625 Hannover, Germany

³Computational Biomolecular Dynamics Group, Max Planck Institute for Biophysical Chemistry, 37077 Göttingen, Germany

⁴School of Engineering, Physics, and Mathematics and Division of Computational Biology, College of Life Sciences, University of Dundee, Dundee DD1 5EH, UK

*Correspondence: jan-philipp.machtens@gmx.de (J.-P.M.), c.fahlke@fz-juelich.de (Ch.F.)

<http://dx.doi.org/10.1016/j.cell.2014.12.035>

SUMMARY

Excitatory amino acid transporters (EAATs) are essential for terminating glutamatergic synaptic transmission. They are not only coupled glutamate/ $\text{Na}^+/\text{H}^+/\text{K}^+$ transporters but also function as anion-selective channels. EAAT anion channels regulate neuronal excitability, and gain-of-function mutations in these proteins result in ataxia and epilepsy. We have combined molecular dynamics simulations with fluorescence spectroscopy of the prokaryotic homolog Glt_{Ph} and patch-clamp recordings of mammalian EAATs to determine how these transporters conduct anions. Whereas outward- and inward-facing Glt_{Ph} conformations are nonconductive, lateral movement of the glutamate transport domain from intermediate transporter conformations results in formation of an anion-selective conduction pathway. Fluorescence quenching of inserted tryptophan residues indicated the entry of anions into this pathway, and mutations of homologous pore-forming residues had analogous effects on Glt_{Ph} simulations and EAAT2/EAAT4 measurements of single-channel currents and anion/cation selectivities. These findings provide a mechanistic framework of how neurotransmitter transporters can operate as anion-selective and ligand-gated ion channels.

INTRODUCTION

Secondary active glutamate transport by excitatory amino acid transporters (EAATs) (Kanner and Sharon, 1978) terminates glutamatergic synaptic transmission and regulates glutamate concentrations within the CNS. EAATs can also function as anion-selective channels (Fairman et al., 1995; Wadiche and Kavanaugh, 1998), with EAAT anion channels regulating cell excitability and synaptic transmission (Picaud et al., 1995). Their physiological relevance is emphasized by the recent discovery that altered EAAT anion conduction is associated with episodic ataxia and epilepsy (Winter et al., 2012).

EAAT anion permeation occurs through a defined anion-selective conduction pathway (Kovermann et al., 2010), which is opened and closed through conformational changes coupled to transitions within the glutamate uptake cycle (Bergles et al., 2002; Machtens et al., 2011a; Otis and Kavanaugh, 2000). The channels are perfectly anion selective (Wadiche and Kavanaugh, 1998) and exhibit unitary current amplitudes, which are small but of a similar size range to those of specialized anion channels (Schneider et al., 2014). The five mammalian EAATs differ in their relative glutamate transport rates and anion currents, resulting in isoform-specific differentiation into efficient transporters associated with small macroscopic anion currents and low-capacity transporters that predominantly conduct anions (Mim et al., 2005). However, the functional properties of the underlying anion channels are very similar for each type (Schneider et al., 2014; Torres-Salazar and Fahlke, 2007), indicating conservation of the anion-conducting pore among functionally specialized transporters. So far, the localization of this conduction pathway, the underlying conformation of the transporter, and the mechanisms of anion permeation have not been identified.

We used molecular dynamics (MD) simulations to identify which conformations of the archeal glutamate transporter homolog Glt_{Ph} (Yernool et al., 2004) permit anion permeation and to characterize the molecular features of anion conduction. We analyzed the conformational changes leading to the formation of an anion-selective pore and observed ion permeation along this path in simulations. Using mutagenesis, fluorescence spectroscopy experiments on Glt_{Ph} and patch-clamp recordings on mammalian EAATs, we confirmed that the anion channel conformation we identified exists under experimental conditions and that this permeation pathway is utilized by both prokaryotic and mammalian glutamate transporters.

RESULTS

Molecular Dynamics Simulations Identify Anion-Conducting Conformations of Glt_{Ph}

Glt_{Ph} shares about 37% sequence identity with mammalian EAATs and is an accepted model of this class of transporters for studying secondary active transport and anion conduction (Boudker et al., 2007; Groeneveld and Slotboom, 2010; Ryan and Mindell, 2007; Yernool et al., 2004). High-resolution crystal

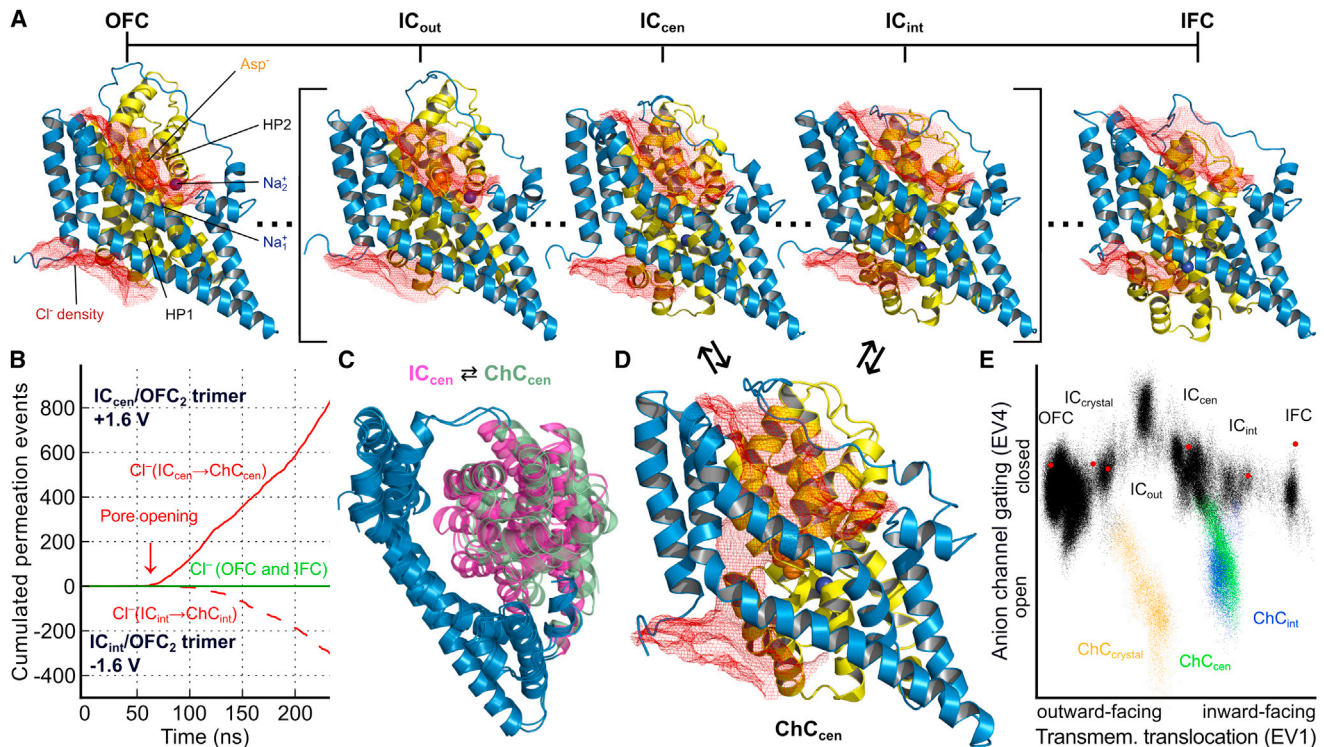


Figure 1. An Anion-Selective Conduction Pathway Is Formed by a Substrate Transport Intermediate

(A) Red isodensity meshes illustrate the Cl⁻ distribution (0.2σ) in MD simulations at +1.6 V around substrate-bound Glt_{ph} monomers in various conformations (trimerization domain shown in blue; transport domain shown in yellow). The other two monomers, water molecules, lipids, and ions were omitted for clarity. Nonconductive outward-facing (OFC), inward-facing (IFC), and intermediate conformations (ICs; derived using essential dynamics sampling of the transition from OFC to IFC) are shown in side view.

(B and C) Ion permeation (B) and conformational change (C) of IC_{cen} during transition to an open channel conformation (overlay of IC_{cen} and ChC_{cen}, top view) upon application of a membrane potential (±1.6 V).

(D) Transitions of ICs into open channel conformations (ChCs) containing an anion-selective pathway occur at positive and negative potentials.

(E) Visualization of all trajectories (the OFC–IFC translocation/essential dynamics sampling simulation and separate MD simulations of OFC, IC_{crystal}, IC_{cen}, IC_{int}, IFC, and of the ChCs of the intermediates) in the principal component space by projection onto the first (EV1) and fourth (EV4) eigenvectors, corresponding to translocation and pore formation, respectively. Black dots represent nonconducting conformations. Blue, green, and orange dots denote frames in MD trajectories, where Cl⁻ permeation through the respective ChC conformation was observed. Note that the point density is biased by the number and length of simulations initiated from the various starting conformations (red circles) and therefore does not provide information on energetics.

See also Figures S1, S2, and S3 and Movie S1.

structures revealed a trimeric assembly, with each subunit containing eight transmembrane helices (TM) and two hairpin loops (HP) (Yernool et al., 2004). Analysis of different conformations demonstrated that substrate translocation involves a large-scale (~18 Å) rotational translational movement of the substrate-harboring transport domain relative to the static trimerization domain (Crisman et al., 2009; Reyes et al., 2009).

We used all-atom MD simulations capable of directly simulating ion flux driven by transmembrane voltages (Kutzner et al., 2011) to investigate anion permeation in substrate-bound Glt_{ph}. Simulations were performed using various Glt_{ph} conformations in the presence of 1 M NaCl on either side of the membrane. Positive and negative membrane potentials (initially ±1.6 V; later ±800 mV) applied to increase anion permeation rates had no detrimental effects on the stability of the system (Figure S1 available online), in good agreement with the results of other simulation studies (Jensen et al., 2012). Within a total simulation time of >8 μs, no Cl⁻ permeation events were observed for the known

outward- (OFC) and inward-facing (IFC) conformations at membrane potentials up to ±1.6 V, indicating that none of these states is anion conducting (Figures 1A and 1B; Extended Experimental Procedures). We concluded that translocation intermediates might correspond to the precursors of anion-conducting conformational states and simulated the OFC–IFC transition to obtain novel intermediate conformations (ICs) using essential dynamics sampling (Amadei et al., 1996). In these simulations, the transporter was driven along the first eigenvector (EV1)—representing transmembrane translocation of the transport domain—from a principal component analysis (PCA) of the conformational changes of transporter monomers in simulations on OFC and IFC (Figures 1A, S1C, and S1D). Because individual subunits function independently within the trimeric assembly (Erkens et al., 2013; Grewer et al., 2005), translocation simulations were performed on a single monomer—with the other two remaining in the OFC. These simulations correctly sampled the recently crystallized Glt_{ph} intermediate (IC_{crystal}, minimum

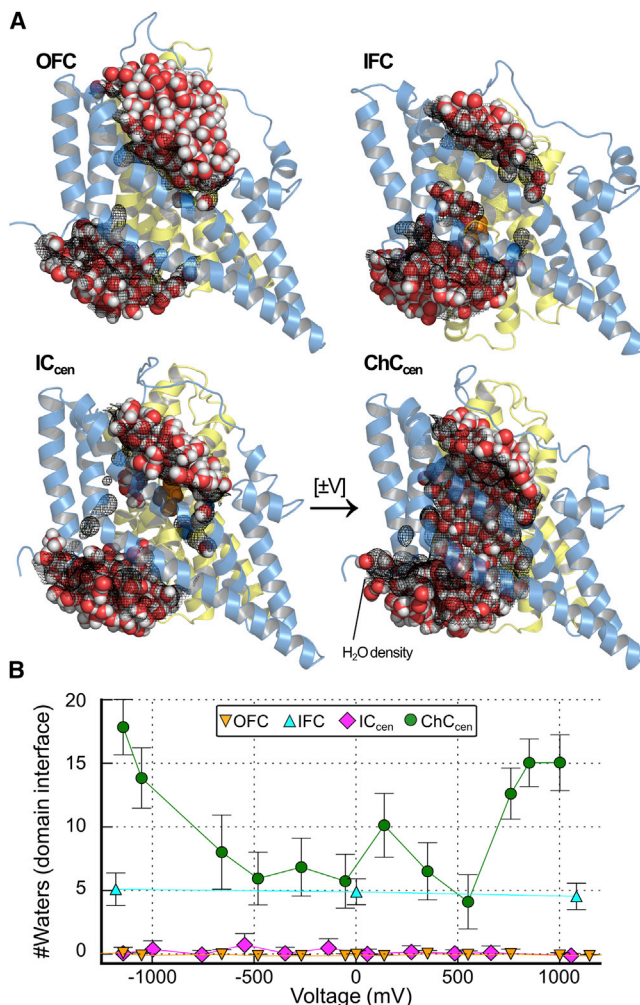


Figure 2. Hydrophobic Gating of the Anion Permeation Pathway

(A) Representative (spheres) and averaged (black mesh) water distribution in the transport/trimerization domain interface of Glt_{Ph} in various conformations (side view).

(B) Voltage-dependent occupancy of the interface core region by water molecules (counted within a cylindrical slab; each data point corresponds to a 100 ns simulation). Note that increased water numbers in ChC, but not in IFC, result in the formation of a continuous water bridge between the extra- and intracellular space (A). Error bars indicate the SD of the water molecule counts during the simulations.

monomeric root-mean-square deviation [RMSD] of 1.3 Å). This demonstrates the existence and stability of translocation intermediates (Verdon and Boudker, 2012) and validates the simulated transition pathway (Figure S2C). We then chose three intermediates, IC_{out} (similar to IC_{crystal}), IC_{cen}, and IC_{int}, from our trajectory that were equally distributed between the OFC and IFC, and, together with IC_{crystal}, subjected them to further MD under transmembrane voltage (Figures 1A and S1).

All intermediates were impermeable to Cl[−] and remained closed for hundreds of nanoseconds in the absence of membrane voltage (Figures 1A and S1G; Extended Experimental Procedures). However, for IC_{crystal}, IC_{cen}, and IC_{int}, lateral movement

of the transport domain occurred 70–300 ns after applying membrane potentials $\geq \pm 1.3$ V. These conformational changes resulted in open channel conformations (designated here as ChC_{crystal}, ChC_{cen}, or ChC_{int}) that were centrally localized on the translocation reaction coordinate and that exhibited an anion-selective conduction pathway at the interface between the trimerization and transport domains, near the tip of HP1 (Figures 1B–1D; Movie S1). Pore opening and closing always reversed after changing the applied voltage, and neither protein instability nor electroporation through the lipid bilayer were observed (Figures S1E–S1H and S3A–S3E). Pore opening occurred from various intermediate conformations, however, with different opening propensities in the order of IC_{out} < IC_{crystal} < IC_{int} < IC_{cen}. For IC_{out} and IC_{crystal}, channel opening was never or only once observed (0 out of 4 [IC_{out}] or 1 out of 5 [IC_{crystal}, at +1.6 V] simulations) within ~400 ns for each (Figures S2D and S2E). In contrast, such transitions were regularly seen for IC_{cen} and IC_{int} (20 out of 20 [IC_{cen}, at 1.3–1.6 V] or 4 out of 4 [IC_{int}, −1.6 V] simulations) (Figure 1B). To further analyze the conformational changes underlying channel opening and to relate them with translocation of Glt_{Ph}, we performed an additional PCA on all data, including the previously used set of OFC and IFC trajectories, the translocation simulations and all simulations of intermediates under membrane voltages. In addition to EV1—which remained unchanged compared with the first PCA and represents translocation—we found that conformational changes along the fourth eigenvector (EV4) correlated with the onset and ending of anion permeation (Figures S3C–S3E). We plotted the position taken up by the trajectories in the principal component space, as defined by eigenvectors EV1 and EV4, which describe conformational changes attributed to translocation and anion channel gating, respectively (Figure 1E). Although originating from different intermediate states, open channel conformations ChC_{cen} and ChC_{int} had RMSDs approaching 1.0 Å with similar overall structures and will be treated as a single conformation (ChC) (Figures 1E and S3F–S3H).

Formation of the anion conduction pore at the interface between the transport and trimerization domains is accompanied by extensive hydration of this region and the creation of a continuous water bridge spanning the membrane (Figure 2A). In each of our simulations, this process was reversible with channel closure preceded by complete dewetting. Water entry is promoted by both positive and negative potentials with voltage-independent water occupancy between −400 and +400 mV (Figure 2B). The hydrophobic environment of this region (see below) is expected to represent a barrier to anion permeation that can be dynamically lowered by the entry of water molecules. Wetting of the rather hydrophobic interface region might compensate for the energetic cost of breaking hydrophobic interactions between the surfaces of the trimerization and transport domains during the conformational change that broadens the interface. Therefore, we suggest that channel opening and closing is mediated by a combination of steric and hydrophobic gating, as has been demonstrated for some other ion channels (Jensen et al., 2012; Vaitheeswaran et al., 2004).

Structural Determinants of Glt_{Ph} Anion Permeation

The Glt_{Ph} anion conduction pathway has a distorted hourglass shape, with large extra- and intracellular entrance cavities that

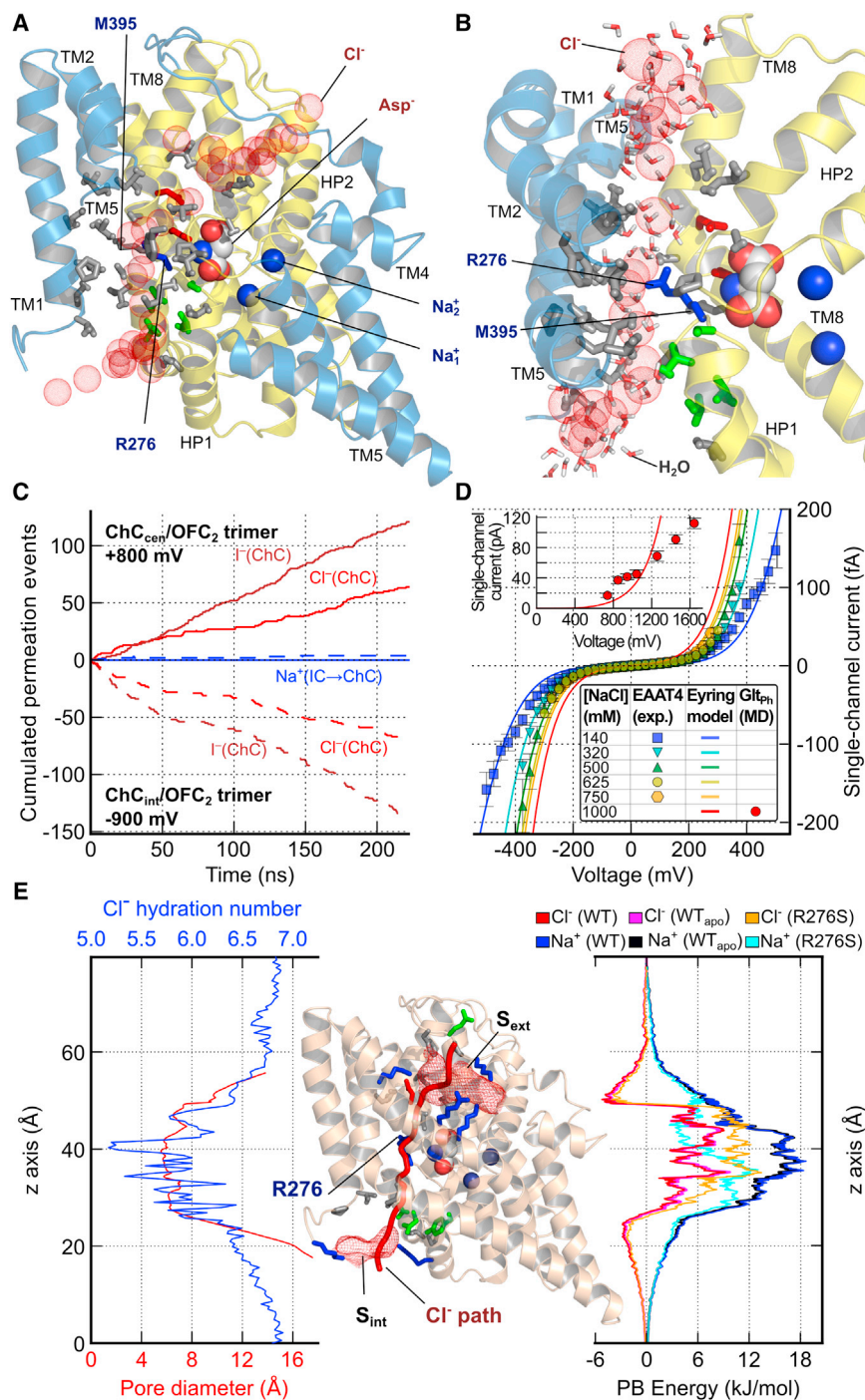


Figure 3. Structural Determinants of Glt_{Ph} Anion Permeation

(A) Cartoon representation of a ChC_{cen} monomer (light blue, trimerization domain; yellow, transport domain) in side view from the subunit interface, with pore-lining side chains shown as sticks (blue, positive; red, negative; green, polar; gray, apolar). TM2 and TM5 are partially omitted for clarity. Red spheres represent snapshots of a single permeating Cl⁻ ion.

(B) Close-up of the permeation pathway from the TM4-TM5 loop. Coloring as in (A), including representative water molecules found in the inner hydration shell of permeating Cl⁻ ions.

(C) Count of Cl⁻ and I⁻ permeation events through ChC_{cen} and ChC_{int} at +800 and -900 mV (dashed lines), respectively.

(D) EAAT4 current voltage plots for various symmetrical [NaCl]. Single-channel currents were determined by multiplying whole-cell Cl⁻ currents recorded 1 ms after the voltage jump (means ± SE, n > 10 for each condition) by the ratio of experimentally measured unitary current (Figure 6B) and mean current amplitudes (n = 12) at +150 mV in 140 mM NO₃⁻. The experimental data (symbols) were globally fitted using a three-binding site Eyring rate model (lines; Extended Experimental Procedures). The inset displays Glt_{Ph} unitary current amplitudes (red symbols) from MD simulations using 1 M NaCl and the extrapolation of the experimental data to these conditions by the Eyring model (red line).

(E) Pore profile of anion hydration, pore diameter, and Poisson-Boltzmann energies for Na⁺ and Cl⁻ of WT and R276S and of WT Glt_{Ph} in an apo state, i.e., after removal of aspartate and Na⁺ ions (in ChC_{cen}). Hydration numbers are the average number of hydrogens within the first Cl⁻ hydration shell. Cl⁻ isodensity meshes around ChC_{cen} (4.2σ) illustrate two Cl⁻-binding sites at the channel entrances, denoted S_{ext} and S_{int}. See also Figure S4.

transport and small anion current components and those with predominant anion conductance (Schneider et al., 2014; Torres-Salazar and Fahlke, 2007). Most side chains lining the pore center are hydrophobic, except for R276, which protrudes from the tip of HP1 into the Cl⁻ density (Figures 3A and 3B). EAATs lack a positive side chain at the position corresponding to R276, but contain arginine at positions homologous to M395 of Glt_{Ph}

narrow to a more constricted conduction path almost perpendicular to the membrane (Figures 1D, 3A, and 3B; Movie S1). Pore-forming residues are highly conserved between Glt_{Ph} and mammalian EAATs (Figure S4). This level of conservation is consistent with the functional similarity between Glt_{Ph} (Ryan and Mindell, 2007) and EAAT anion channels (Melzer et al., 2003; Wadiche and Kavanaugh, 1998) and accounts for similar unitary current amplitudes among EAATs with large glutamate

(Ryan et al., 2010) (Figure S4). MD simulations of the R276S-M395R Glt_{Ph} mutant (see below) showed that both the arginine in the “EAAT position” and R276 project their side chain toward the same location, resulting in conservation of the positive charge at this site in the tertiary structure of EAATs and Glt_{Ph} (Figures 3A and 3B).

Starting from the ChC_{cen} conformation, we simulated anion permeation at reduced voltages of ±~800 mV and observed

perfect anion selectivity (Figure 3C). In 1 M NaCl, our simulations yielded single-channel anion currents of 42.4 ± 6.3 pA ($\sim +800$ mV) or 51.4 ± 6.7 pA (~ -900 mV). These voltages and salt concentration are too high to permit direct experimental verification of the conductances. We measured EAAT4 anion currents at 140–750 mM NaCl and at voltages up to 500 mV (Figure 3D) to extrapolate the voltage dependence of EAAT4 unitary anion currents to the MD conditions. Comparing these experimental EAAT4 and simulated Glt_{Ph} current-voltage relationships demonstrated that simulations reproduce the experimental unitary current amplitudes within the same order of magnitude (Figure 3D, inset). Substitution of Cl[−] by I[−] in the simulations resulted in significantly higher anion currents of 95.0 ± 5.4 pA ($\sim +800$ mV) or of 97.1 ± 10 pA (~ -900 mV; Figure 3C); however, the transport substrate aspartate did not permeate within 200 ns at concentrations of ~ 500 mM. Simulated permeation properties thus closely resemble the functional characteristics of mammalian EAATs (Melzer et al., 2003; Wadiche and Kavanaugh, 1998).

The electrostatic Poisson-Boltzmann energy profile for moving an ion along the channel axis displays much higher energy barriers for Na⁺ than for Cl[−] (Figure 3E). Energy wells at both entrances with high Cl[−] densities represent Cl[−] binding sites, denoted S_{ext} and S_{int}. The critical role of R276 in anion selectivity is demonstrated by the convergence of Na⁺ and Cl[−] energy barriers upon removal of the positive charge in R276S Glt_{Ph} (Figure 3E). Energy profiles are identical in both the presence and absence of bound aspartate/Na⁺ at their binding sites (Figure 3E). Simulated ion permeation through this apo state revealed similar Cl[−] permeation rates along the same permeation pathway (data not shown), consistent with the experimentally determined unitary conductances of EAATs being indistinguishable in the presence and absence of substrate (Kovermann et al., 2010). The conduction pathway is rather wide with a minimum diameter of 5.6 Å, such that anions can permeate in a partially hydrated state, and Cl[−]-H₂O coordination numbers show only a small decrease from 6.8 in bulk solution to 5.2 in the Glt_{Ph} channel center (Figure 3E).

Tryptophan-Scanning Mutagenesis Reveals Direct Interactions of Predicted Pore-Forming Residues with Permeant Anions

We used a combination of tryptophan-scanning mutagenesis and iodide quenching (Vázquez-Ibar et al., 2004) to test whether permeating anions come into close contact with amino acid side chains projecting into the proposed anion conduction pathway. I[−] readily permeates through Glt_{Ph} and EAAT anion channels and is therefore expected to come into close proximity to residues forming the permeation pathway. Because I[−] can reduce tryptophan fluorescence via direct interactions, i.e., collisional quenching (Lakowicz, 2006), iodide quenching of tryptophan fluorescence is a suitable method to experimentally verify the simulated Glt_{Ph} anion permeation pathway. As Glt_{Ph} lacks endogenous tryptophans, we generated single-tryptophan mutants by substituting 13 residues that protrude from the trimerization domain into the interface region of Glt_{Ph} ChC_{cen} (Figure 4A). To avoid interference with substrate binding, we did not insert tryptophan residues into the transport domain.

With the exception of S65W, I[−] reduced fluorescence in all Glt_{Ph} mutants in a concentration-dependent manner. Figure 4B shows the spectral properties of V51W and S65W Glt_{Ph} in detergent micelles and their modification at various [I[−]]. The identical concentration dependences of the fluorescence lifetimes and intensities indicate that I[−] quenches tryptophan fluorescence via a collisional mechanism (Figure 4B, inset).

Figure 4A maps the relative quenching (F_0/F) at [I[−]] = 350 mM of the tested tryptophan residues on the ChC_{cen} structure. These data demonstrate the high iodide accessibility of residues close to the proposed anion permeation pathway, which is reduced with increasing distance. Linear concentration dependences of F_0/F in Stern-Volmer plots are expected for proteins with a single tryptophan, which adopt only one conformation. The observed deviations from linearity indicate that tryptophan-substituted Glt_{Ph} mutants assume multiple conformations that differ in I[−] accessibility (Figure 4C). These findings support the notion that tryptophan-substituted Glt_{Ph} mutants exhibit similar degrees of conformational heterogeneity to wild-type (WT) Glt_{Ph}.

Figure 4D shows plots of the calculated anion accessibilities in simulations of different conformations compared with experimentally observed fluorescence quenching. Most residues are accessible in multiple conformations. However, there are three residues—W50, W54 (projecting directly into the anion pore), and W62 (at a more peripheral location)—with rather exclusive anion accessibility in ChC (Figures 4D and S5; see the Extended Experimental Procedures for details on the calculation of anion accessibilities from the simulations). For these constructs, a modified Stern-Volmer analysis was used to determine the fraction of fluorescence quenchable by I[−] to be $\sim 20\%$ (Figures S5A and S5B). Because different protein conformations could exhibit different quantum yields of the inserted tryptophan, this value is not always identical to the probability of the protein assuming this accessible conformation. However, these data indicate that Glt_{Ph} can assume the anion-conducting channel conformation ChC even in the absence of an applied voltage and that this conformation is sufficiently stable to permit I[−] collisions with side chains that project into the anion conduction pathway.

Mutations of Pore-Forming Residues Affect Unitary Anion Current Amplitudes and Anion/Cation Permeability Ratios

To provide further verification of the predicted anion permeation pathway, we compared the effects of amino acid exchanges on simulated and experimental permeation properties. We chose experimental measures corresponding to parameters obtained from MD simulations: these included the single-channel conductance and the anion/cation selectivity of the anion channel. In contrast, macroscopic current amplitudes alone, e.g., from whole-cell recordings, do not permit a distinction to be made between mutations that alter the anion permeation rate or those that affect the probability of assuming an open anion channel state and therefore preclude a direct comparison with simulation results. Because of difficulties in cellular expression systems, high-resolution electrophysiological recordings of Glt_{Ph} are not yet feasible. Assays that were developed to describe Glt_{Ph} anion conductance (Ryan and Mindell, 2007) only provide information

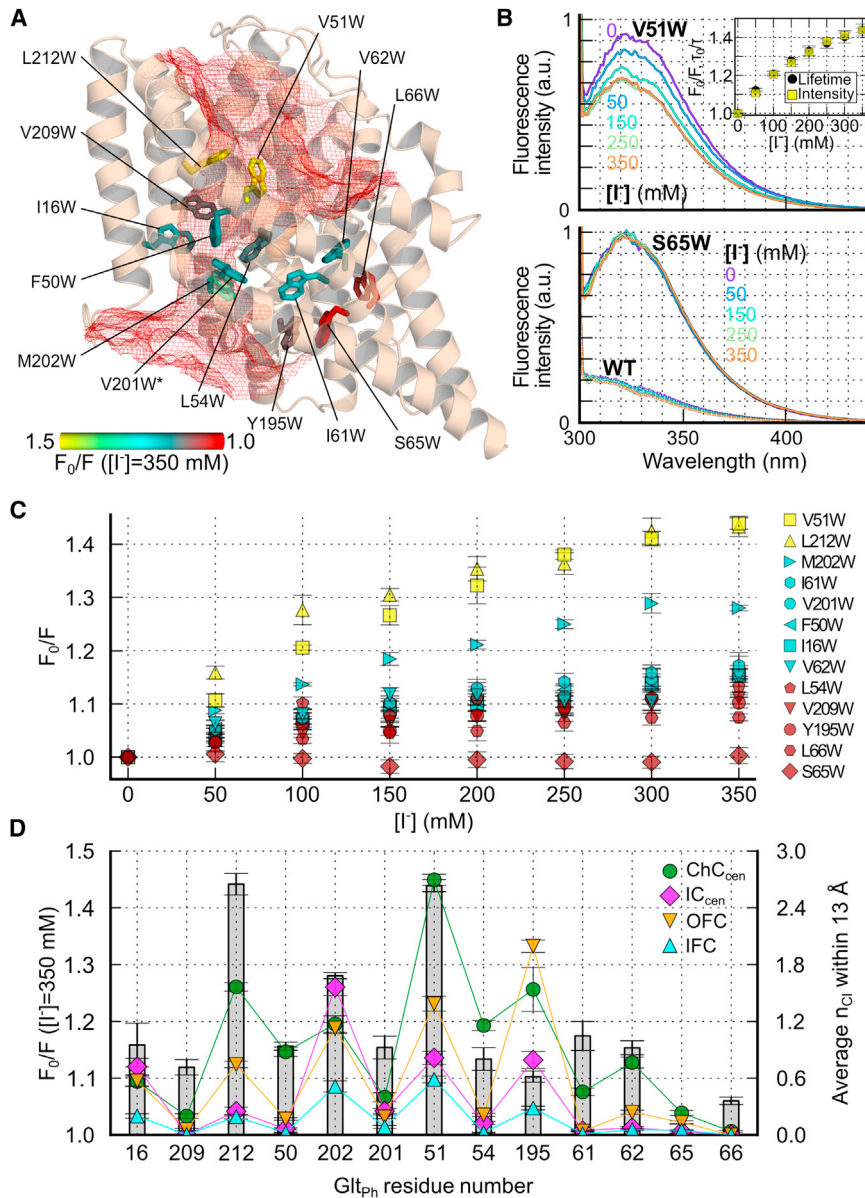


Figure 4. Tryptophan Fluorescence Quenching by Iodide in Glt_{P_H}

(A) Overview of Glt_{P_H} single-tryptophan insertions (ChC_{cen} in side view). Side chains are color-coded according to the reduction in fluorescence intensity at 350 mM [I⁻] (F₀ is the intensity in the absence of I⁻; n ≥ 5 for each). Red mesh represents the Cl⁻ density observed in MD (Figure 1D). (B) Representative fluorescence spectra of WT, V51W, and S65W Glt_{P_H} at various [I⁻]. The inset shows the comparable concentration dependence of V51W fluorescence lifetimes and intensities in a Stern-Volmer plot, indicating a collisional quenching mechanism. (C) Stern-Volmer plots for all tryptophan mutants (means ± SE; n ≥ 5 for each). (D) Comparison of fluorescence quenching (gray bars; n ≥ 5 for each) with MD-predicted anion accessibilities of side chains in various conformations (the different symbols show the average number of Cl⁻ ions ± SD within 13 Å of the side chains). Residue numbers on the abscissa are ordered according to their positions in the membrane plane shown in (A). See also Figure S5.

We initially screened for mutations that affect pore properties in silico. Pore-forming residues were identified by a geometrical criterion (within a distance <6.9 Å to the pore center defined by R276). We excluded only a few residues that are located within the transport domain and known to be crucial for substrate binding (e.g., residues in HP2 to prevent interference with substrate binding). We generated 29 Glt_{P_H} pore mutants in silico, including S65 and I61, which are close to a recently discussed alternative location of the anion channel (see below) and subjected them to MD simulations. We identified side-chain substitutions that increase or decrease unitary conductances (I16E/K, L20E, F50L/K/D, V51D, L54D, I61D,

about macroscopic anion currents through an ensemble of multiple Glt_{P_H} transporters. Because the functional properties of Glt_{P_H} (Ryan and Mindell, 2007) and EAAT anion channels are very similar (Melzer et al., 2003; Wadiche and Kavanaugh, 1998) and the pore-lining residues are highly conserved (Figure S4), it is reasonable to assume that the proposed Glt_{P_H} anion permeation pathway is also responsible for EAAT anion conduction. We therefore compared the effects of in silico mutagenesis on simulated Glt_{P_H} anion conductance and anion/cation selectivity with experimental data on mammalian EAAT2/EAAT4. Single-channel recordings have not yet been possible for these transporters, but unitary current amplitudes can be determined by noise analysis of whole-cell current recordings, and anion/cation selectivities can be obtained through reversal potential measurements at various ionic conditions (Melzer et al., 2003).

A205D, R276S/D) or modify the anion/cation selectivity (F50D, A205D, R276S) (Figure 5). We then performed whole-cell patch-clamp recordings of 33 EAAT2/EAAT4 mutants (Figure S6). Because most mutations also affected anion channel gating (Figure S6), a direct comparison of whole-cell currents and MD data was not feasible. However, ten EAAT4 mutants exhibited sufficient time- and voltage-dependent gating to allow single-channel conductances to be determined using nonstationary noise analysis (Figure 6). Four charge-altering mutations, L20E, I16K, I16E and I61D, increased or decreased the simulated Cl⁻ permeation rate of Glt_{P_H} to a similar extent as alterations in experimental single-channel conductance caused by the homologous EAAT4 mutations. For four Glt_{P_H} mutants, V12E, I16W, and S65A, experimental and simulated unitary conductances were unaltered (Figure 6B). Interestingly, I16E—located in the intracellular part of the

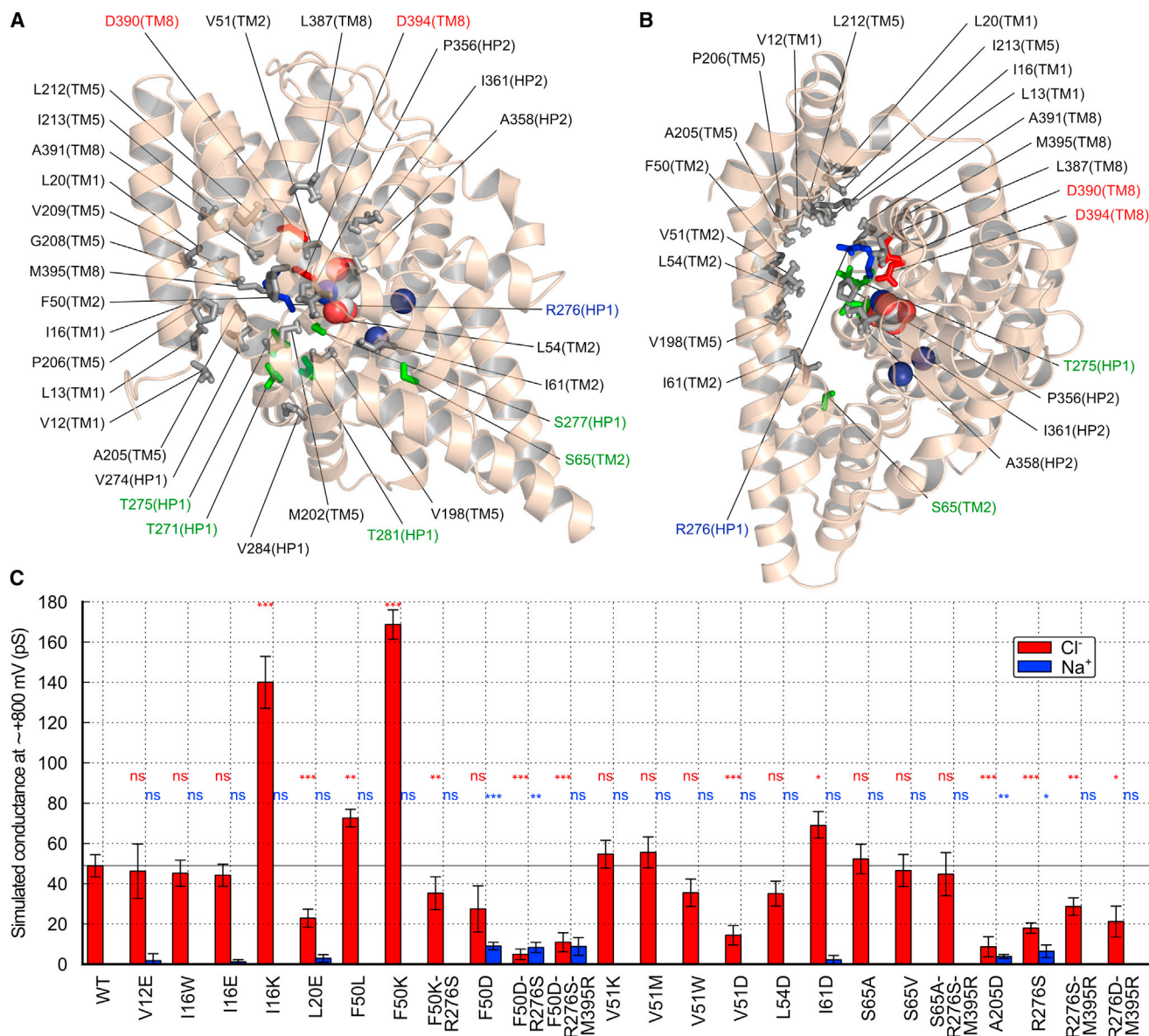


Figure 5. MD Screening of Pore-Lining Residues Predicts Mutations that Affect Anion Conductance and Anion/Cation Selectivity

(A and B) Stick representations of pore-lining residues in side view (A) or top view (B) and colored as in Figure 3A, including detailed GltpH residue number labels. (C) Summary of simulated Na^+/Cl^- conductances for various GltpH mutants (ChC_{gen} ; means \pm SD; MD times range from 120 to 500 ns for each mutant) at $\sim +800$ mV.

GltpH anion conduction pathway—selectively reduced outward fluxes of anions in a valve-like manner, as demonstrated by outward current rectification in the I16E GltpH and homologous T59E EAAT4 mutants. In contrast, for the neighboring residue V12, which is closer to the bulk solution and further from the pore center than I16 (Figure 5A), conversion to glutamate did not affect unitary conductances in GltpH or in F55E EAAT4. We furthermore found four “semiconserved” side chains that are conserved within mammalian EAATs but differ in GltpH : the aforementioned R276 residue (the corresponding EAAT arginine is located at the M395 position in GltpH), F50 (L in EAATs), and M94 (V in

EAATs; Figure S4). We constructed GltpH and EAAT4 mutants to reverse these evolutionary exchanges and observed reciprocal effects on conductance, as would be expected if direct interactions exist between these side chains and permeating anions (Figure 6C).

The simulated GltpH mutants F50D, A205D, and R276S exhibited Na^+ permeation along the same path as Cl^- (Figures 7A and 7C). Because some of the corresponding EAAT4 mutations prevented their functional expression in cells, homologous mutations were introduced into EAAT2, a transporter with unitary current properties similar to those of EAAT4 (Schneider et al., 2014).

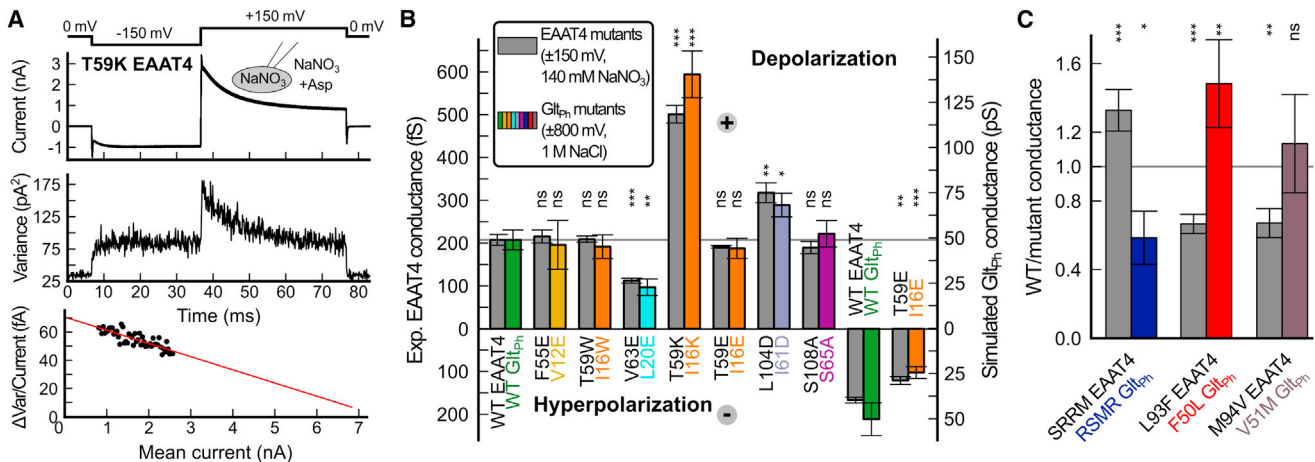


Figure 6. Mutations of Pore-Forming Residues Modify Experimental EAAT4 Anion Conductances and Anion/Cation Selectivity

(A) Representative nonstationary noise analysis of T59K EAAT4, showing current responses to 300 repeated voltage jumps (top) and the resulting current variances (middle). Bottom, linearly transformed current-variance plot (background noise at 0 mV was subtracted from the total variance). Red line shows a linear fit. (B) Experimental EAAT4 (gray; from whole-cell recordings and nonstationary noise analysis; means \pm SE; $n = 6-9$) and simulated Glt_{ph} unitary conductances of WT and homologous mutants (in color; means \pm SD). Ordinates were scaled to show experimental WT EAAT4 conductances at +150 mV and simulated WT Glt_{ph} conductances at +800 mV at the same level (gray line). (C) Changes in experimental EAAT4 ($n = 6-8$) and simulated Glt_{ph} unitary conductance upon substitution of residues that are conserved in EAATs, but not in Glt_{ph} (Figure S4).

See also Figures S6 and S7 and Table S1.

Varying the external $[Na^+]$ led to changes in ion current reversal potentials in cells expressing L85D, S288D, and R476M EAAT2, indicating that EAAT2 mutants represent unselective channels with varying degrees of relative cation selectivity (Figure 7B). In these experiments, coupled glutamate transport, which would additionally affect the reversal potential, was abolished by using a K^+ -free intracellular solution.

The effect of these negative charge mutations was site-specific as demonstrated by experiments and simulations with A362D EAAT2 and the corresponding R276D-M395R Glt_{ph}. A362 in EAAT2 is homologous to R276 in Glt_{ph}, whose positively charged side chain is crucial for anion selectivity (Figure 3). To achieve the electrostatic potential in Glt_{ph} to be similarly modified as in A362 EAAT2, we inserted an arginine at the “EAAT position” M395, in addition to the R276D mutation (Figures 3A and S4). R276D-M395R Glt_{ph} and A362D EAAT2 exhibited perfect Cl^- selectivity in both simulations and experiments, indicating that anion selectivity is only impaired by the insertion of negatively charged side chains at specific positions (Figures 5 and 7).

The Novel Anion Channel Conformation Enables a Reinterpretation of Previous Functional Data and Is Compatible with Published Crosslinking Results

Prior to our work, the structural basis of EAAT/Glt_{ph} anion conduction was unknown. However, because mutations around S65 were reported to affect anion permeation of both Glt_{ph} (Ryan and Mindell, 2007) and EAAT1 (Cater et al., 2014; Ryan et al., 2004), and because crystallographic data (Verdon and Boudker, 2012) suggested the existence of an aqueous cavity in IC_{crystal}, it has been hypothesized that anions permeate along a pathway that we will refer to as “S65 path” (Figure S7A). As yet, no other EAAT anion permeation pathway has been proposed.

We performed MD simulations and experimental approaches to test whether anion permeation along the “S65 path” might contribute to EAAT/Glt_{ph} anion conduction. MD simulations demonstrated water access but no Cl^- density along the “S65 path” in ChC (Figure S7A). Pore searching algorithms (see the Extended Experimental Procedures) did not identify any additional candidate anion pore in the S65 region. Mutations of S65 did not affect anion conductance in MD simulations (Figure 5). In our fluorescence assay, S65W Glt_{ph} fluorescence was not quenched by iodide (Figure 4). Whereas the homologous S108V EAAT4 mutant was mostly retained in intracellular compartments, S108A EAAT4 was robustly expressed on the surface of mammalian cells, with resulting current amplitudes comparable with those of WT EAAT4. S108A EAAT4 exhibited altered anion channel gating but unaltered unitary current amplitudes (Figures 6B and S6). These results indicate that mutations of S65/S108 do not affect the single-channel conductance itself but instead alter the channel open probability, i.e., the rates of reactions leading to the open anion channel.

We generated three additional EAAT4 mutants (V101D, L104D, and N297D) with negatively charged side chains projecting into the “S65 path.” Mutant channels exhibited altered voltage- and glutamate-dependent gating but were still glutamate sensitive and cation impermeable (Figures S6B, S6C, S7A, and S7B). One mutation in this region, I61D Glt_{ph}/L104D EAAT4, even increased anion permeation rates in both simulations and experiments (Figures 5 and 6B). Because this residue does not directly line the Cl^- permeation pathway, which remained unchanged upon I61D substitution, and because the introduction of a negative charge increases anion conductance, we deduce that this mutation indirectly affects the anion channel function. We conclude that the mutated amino acids surrounding

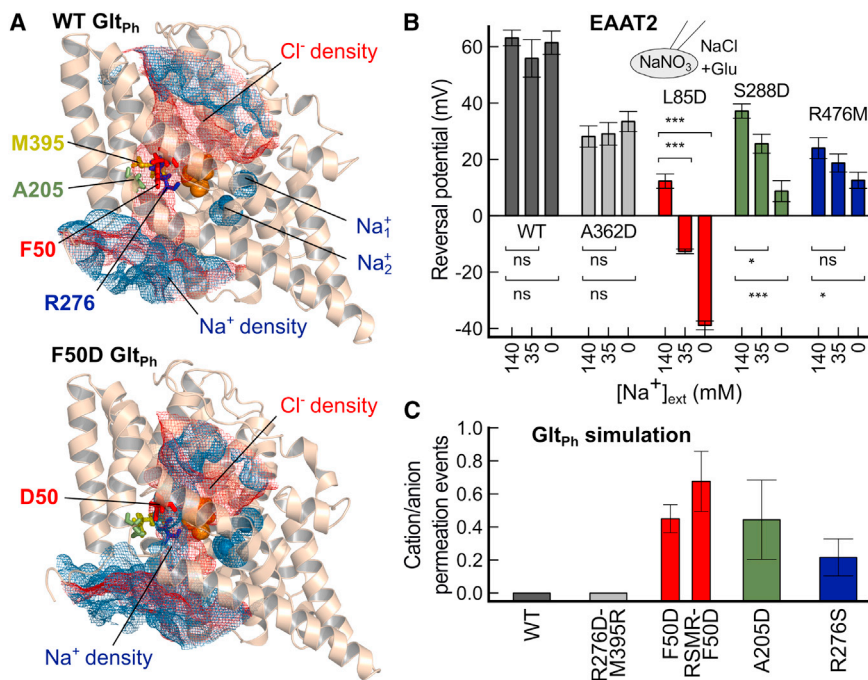


Figure 7. Conversion of the EAAT2 Anion Pore into a Cation-Conducting Channel

(A) Cl⁻ (red) and Na⁺ (blue) distributions ($\sigma = 0.2$) around WT and F50D Glt_{ph}. Residues described in (B) and (C) are shown as sticks.

(B) Variations in current reversal potentials with external [Na⁺] demonstrate the cation permeability of L85D EAAT2—homologous to F50D Glt_{ph}—S288D, and R476M (but not of WT and A362D) EAAT2 anion channels (means \pm SE; $n = 6$ –13 for each).

(C) Ratio of simulated cation/anion permeation events for WT and corresponding Glt_{ph} mutants (means \pm SD; RSMR, R276S-M395R), colored according to (B). The F50D mutant was tested with the arginine at position 276 (WT) and in the context of the EAAT arginine position (RSMR mutation). See also Figures S6 and S7 and Table S1.

DISCUSSION

EAAT glutamate transporters are prototypical dual function proteins that operate as both secondary active transporters and anion-selective ion channels.

Whereas the key structural features of secondary active glutamate transport have been established (Akyuz et al., 2013; Crisman et al., 2009; Reyes et al., 2009; Shrivastava et al., 2008), structural and mechanistic details of anion permeation have been hitherto unknown. In this study, we used a combination of computational and experimental approaches to determine how this class of transporters mediates anion permeation through an aqueous conduction pathway. MD simulations identified an open channel conformation of Glt_{ph} that was consistently formed from various ICs by the lateral movement of the transport domain (Figure 1). Opening of the interface between the transport and trimerization domains is followed by voltage-promoted water entry (Figure 2) and the formation of an anion-selective conduction pathway (Figure 3). We verified the predictions of our simulations by fluorescence spectroscopy and functional studies using mutant transporters. Fluorescence quenching experiments demonstrated that tryptophan residues substituted at positions that project into the predicted conduction pathway come into close contact with permeating anions (Figure 4). Moreover, substitution of pore-forming residues had comparable experimental effects on the two key characteristics of an anion-selective conduction pathway, i.e., anion/cation selectivity and ion permeation rates, as predicted by simulations (Figures 5, 6, and 7). These data indicate that pore-forming residues identified through simulations are indeed the major determinants of anion permeation and selectivity in both Glt_{ph} and EAATs. Moreover, they demonstrate that this anion conduction pathway is conserved throughout the glutamate transporter family. Our data thus clarify how a class of secondary active transporters can function as anion-selective channels that are gated by transitions in the transport cycle.

The ion conduction pathway reported herein accounts for all known functional properties of EAAT/Glt_{ph} anion channels.

S65 in Glt_{ph} do not line the EAAT anion pore, although they do influence the conformational changes underlying the probability of the channel being open.

A recent study demonstrated that crosslinking a substituted cysteine within the transport domain to another in the trimerization domain abolishes EAAT3 glutamate transport but does not abrogate substrate-dependent anion conductance (Shabaneh et al., 2014). The authors concluded that, starting from OFC, a limited inward movement of the transport domain is sufficient for formation of an anion conducting conformation. Cysteines were inserted at positions corresponding to residues 216 and 391 in Glt_{ph}. These residues are in close proximity in OFC, IC_{out}, and IC_{crystal}. Because MD simulations demonstrated a pronounced increase in the V216–A391 C _{α} distance to >7 Å during transmembrane translocation and channel opening (Figure S7C), this disulfide link might prevent transitions into the IC_{cen}, IC_{int}, ChC, and IFC states. To evaluate the effects of this disulfide bridge on the conformational changes underlying anion channel opening, we performed simulations on an intermediate conformation of our translocation trajectory that is located at the most central position along the translocation axis (to increase the likelihood of pore opening) but maintains a distance between these two residues of ≤ 7 Å (Figures 1 and S7C). The crosslinkage was modeled by a distance restraint on the two C _{α} atoms within monomers (Figures S7C and S7D). Simulations of the V216–A391 crosslinked Glt_{ph} model showed that this disulfide link limits the lateral movement of the transport domain but permits sufficient conformational flexibility for pore opening and anion permeation along the identified anion conduction pore (Figures S7E and S7F). The experimental effects of this crosslink on transport and anion currents in EAAT3 (Shabaneh et al., 2014) are therefore fully consistent with the Glt_{ph} anion permeation pathway presented here.

Simulations reveal unitary current amplitudes and ion selectivities (Figures 3C and 3D) that resemble experimental results (Melzer et al., 2003; Wadiche and Kavanaugh, 1998). The calculated minimum pore diameter of Gltp_h is ~5.6 Å (Figure 3E), which perfectly fits the predicted minimum pore diameter of >5 Å based on anion substitution experiments on EAAT1 anion channels (Wadiche and Kavanaugh, 1998). Rapid substrate application experiments have shown that EAAT anion channel activation is delayed compared with glutamate translocation (Grewer et al., 2000). These findings indicate that anion-conducting states existing “outside” the glutamate uptake cycle can be explained by channel opening as a branching reaction from ICs (Figure 1). Simulations predict voltage independence of anion channel opening within the physiological voltage range (Figure 2B). This result explains the experimental observation that the voltage- and substrate dependence of EAAT anion channels are tightly linked to transitions within the transport cycle (Bergles et al., 2002; Machtens et al., 2011a). Simulated anion permeation is unchanged in both the presence and absence of bound substrate (Figure 3E), as expected from the experimental unitary conductances being indistinguishable in the presence and absence of glutamate (Kovermann et al., 2010). Because anion channel opening is tightly linked to translocation of the transport domain, our results indicate that transport substrates increase EAAT anion currents by promoting intermediate states. Distinct EAAT isoforms differ strongly in the relative amplitudes of their transport and anion currents (Fairman et al., 1995; Mim et al., 2005). However, analysis of unitary current amplitudes revealed similar single-channel amplitudes (Schneider et al., 2014; Torres-Salazar and Fahlke, 2007). The high degree of conservation of pore-forming residues (Figure S4) is consistent with the similarities in anion channel unitary current amplitudes and selectivity of different transporter isoforms. Lastly, the novel anion conducting conformation can account for all published mutagenesis and crosslinking results on EAAT anion conduction (Ryan et al., 2004; Shabaneh et al., 2014).

The “S65 path” (Figure S7) is the only location of the anion channel that has been discussed in recent years. We could not find any indication for a direct contribution of this region to anion permeation. Our simulations show that the “S65 path” is hydrated in ChC, thereby suggesting that S65 and adjacent residues could be involved in facilitating the opening of the transport/trimerization domain interface instead. We thus speculate that the “S65 path” may modulate formation of the ChC conformation, which provides an explanation for the impact of mutations in this region on anion channel function (Cater et al., 2014; Ryan and Mindell, 2007; Ryan et al., 2004).

The positive electrostatic potential necessary for perfect anion selectivity of EAAT/Glt_{ph} anion channels is provided by a single positively charged side chain, R276. Surprisingly, during evolution, this arginine has moved from the tip of HP1 in Glt_{ph} to TM8 in EAATs, while retaining a similar side chain position in the tertiary structure. In Glt_{ph}, as well as in EAATs, this arginine has been implicated in binding amino acid substrates, as well as binding Na⁺ and K⁺ (Ryan et al., 2010; Verdon et al., 2014). Unitary anion conductance is not affected by aspartate (Figure 3), indicating that the interaction of R276 with transport substrates does not modify its effect on anion conduction and selectivity.

The tight linkage between anion channel gating and glutamate transport in EAAT/Glt_{ph} was previously explained by assuming that certain states of the transport cycle are anion conducting (Bergles et al., 2002). Because Glt_{ph} structures did not exhibit an open pore with dimensions that might account for the experimentally observed anion conduction properties, it was recently suggested that additional yet to be defined ICs that occur during translocation might be anion conducting (Cater et al., 2014). We have now demonstrated that intermediate transport conformations are nonconducting and that EAAT/Glt_{ph} anion channel opening transitions require the lateral movement of the glutamate transport domain together with pore hydration from intermediates. Anion channel opening is therefore not part of the transport cycle, but instead is achieved via a branching conformational change. This design permits rapid transition through the full transport cycle without anion channel opening. Furthermore, it allows certain EAAT isoforms to function as effective transporters, with low anion channel open probabilities, and other isoforms to have low transport rates but high occupations of the anion channel mode.

The unique mechanism of EAAT anion channel gating results in neuronal or glial anion conductances that follow changes in substrate concentrations and thus allow feedback control of glutamate release (Wersinger et al., 2006) or modification of GABAergic postsynaptic currents by glutamatergic signals (Winter et al., 2012). Moreover, it explains why isoform-specific variations in glutamate transport by EAATs result in the formation of anion channels that preferentially open or close within their physiological voltage range (Schneider et al., 2014). Recently, gain-of-function mutations in genes encoding EAAT anion channels have been linked to pathological neuronal excitability and cell-volume regulation (Winter et al., 2012). EAAT anion channel activity is also enhanced under conditions of increased synaptic glutamate concentration and may thus contribute to the clinical symptoms associated with brain ischemia or certain neurodegenerative diseases. The structural and mechanistic data presented here might help in the design of EAAT anion channel modulators and thus open therapeutic avenues to correct the cellular defects linked to these pathological conditions.

EXPERIMENTAL PROCEDURES

Molecular Simulations

MD simulations of Glt_{ph}—bound by a negatively charged aspartate and two Na⁺ ions—in outward-facing (OFC; Protein Data Bank [PDB] ID code 2NWX), inward-facing (IFC; PDB ID code 3KBC), and various intermediate conformations (including IC_{crystal}; PDB ID code 3V8G) were performed using GROMACS 4.5 (Hess et al., 2008). Based on our OFC and IFC simulation trajectories, we obtained intermediates IC_{int}, IC_{cen}, and IC_{out} from the crystallographic structures of OFC and IFC using essential dynamics sampling simulations (Amadei et al., 1996). Proteins were inserted and equilibrated in a double dimyristoyl phosphatidylcholine bilayer surrounded by a 1 M NaCl aqueous solution and were subjected to various membrane potentials using the computational electrophysiology scheme described recently (Kutzner et al., 2011).

Molecular Biology

Mutant constructs of Glt_{ph}, human EAAT4, and rat EAAT2 were generated using the QuikChange Site-Directed Mutagenesis Kit (Agilent Technologies) and verified by restriction analysis and DNA sequencing.

Fluorescence Spectroscopy

Fluorescence emission spectra of single-tryptophan GlT_{PH} mutants in n-dodecyl-β-D-maltoside micelles in the presence of saturating [Na⁺] and [Asp⁻] at various [I⁻] were recorded after excitation at 295 nm. Fluorescence lifetimes were determined through time-correlated single-photon counting.

Electrophysiology

Heterologous expression and whole-cell patch-clamp recordings of EAAT2 and EAAT4 were performed as described previously (Machtens et al., 2011a). Unitary conductances were determined by nonstationary noise analysis of current responses to 300 repetitive voltage jumps to ±150 mV using 140 mM NO₃⁻ as main anion and 1 mM aspartate as substrate to enhance voltage-dependent gating of the channel.

Statistics

Asterisks indicate the level of statistical significance derived from a two-tailed t test (**p < 0.001; *p < 0.01; *p < 0.05; ns, p ≥ 0.05; see Table S1).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, one table, and one movie and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.12.035>.

AUTHOR CONTRIBUTIONS

J.-P.M. and Ch.F. designed the research; J.-P.M. carried out MD simulations; J.-P.M., P.K., and B.B. generated mutant DNA constructs; J.-P.M., D.K., and A.L. performed patch-clamp recordings; C.L. and D.E. conducted fluorescence spectroscopy measurements; J.-P.M. and D.E. analyzed data; B.L.d.G., R.B., and U.Z. advised on the setup and analysis of MD simulations; and J.-P.M., D.E., R.B., B.L.d.G., and Ch.F. prepared the figures and wrote the paper.

ACKNOWLEDGMENTS

These studies were supported by the Deutsche Forschungsgemeinschaft (FA301/9 to Ch.F.; SFB803 to B.L.d.G. and R.B.). The authors gratefully acknowledge the computing time granted on the supercomputer JUROPA at Jülich Supercomputing Centre (JSC) and on the HLRN-II supercomputer of the North-German Supercomputing Alliance (HLRN). We thank A. Alekov, K. Benndorf, H. Grubmüller, and P. Hidalgo for critically reading the manuscript; T. Becher, R.E. Guzman, P. Kovermann, G. Stölting, B. Wilhelm, T. Gensch, I. Weyand, D. Köpfer, D. Wojciechowski, and D.J. Slotboom for helpful discussions; and T. Wassmann for preliminary simulations.

Received: May 23, 2014

Revised: August 27, 2014

Accepted: December 19, 2014

Published: January 29, 2015

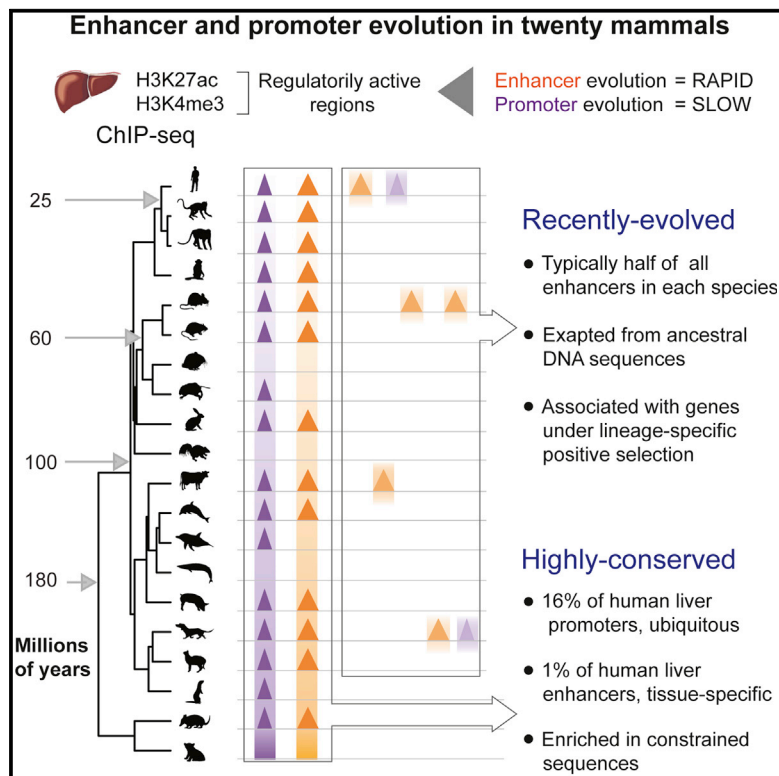
REFERENCES

- Akyuz, N., Altman, R.B., Blanchard, S.C., and Boudker, O. (2013). Transport dynamics in a glutamate transporter homologue. *Nature* 502, 114–118.
- Amadei, A., Linssen, A.B.M., de Groot, B.L., van Aalten, D.M.F., and Berendsen, H.J.C. (1996). An efficient method for sampling the essential subspace of proteins. *J. Biomol. Struct. Dyn.* 13, 615–625.
- Bergles, D.E., Tzingounis, A.V., and Jahr, C.E. (2002). Comparison of coupled and uncoupled currents during glutamate uptake by GLT-1 transporters. *J. Neurosci.* 22, 10153–10162.
- Boudker, O., Ryan, R.M., Yernool, D., Shimamoto, K., and Gouaux, E. (2007). Coupling substrate and ion binding to extracellular gate of a sodium-dependent aspartate transporter. *Nature* 445, 387–393.
- Cater, R.J., Vandenberg, R.J., and Ryan, R.M. (2014). The domain interface of the human glutamate transporter EAAT1 mediates chloride permeation. *Biophys. J.* 107, 621–629.
- Crisman, T.J., Qu, S., Kanner, B.I., and Forrest, L.R. (2009). Inward-facing conformation of glutamate transporters as revealed by their inverted-topology structural repeats. *Proc. Natl. Acad. Sci. USA* 106, 20752–20757.
- Erkens, G.B., Hänelt, I., Goudsmits, J.M., Slotboom, D.J., and van Oijen, A.M. (2013). Unsynchronised subunit motion in single trimeric sodium-coupled aspartate transporters. *Nature* 502, 119–123.
- Fairman, W.A., Vandenberg, R.J., Arriza, J.L., Kavanaugh, M.P., and Amara, S.G. (1995). An excitatory amino-acid transporter with properties of a ligand-gated chloride channel. *Nature* 375, 599–603.
- Grewer, C., Watzke, N., Wiessner, M., and Rauen, T. (2000). Glutamate translocation of the neuronal glutamate transporter EAAC1 occurs within milliseconds. *Proc. Natl. Acad. Sci. USA* 97, 9706–9711.
- Grewer, C., Balani, P., Weidenfeller, C., Bartusel, T., Tao, Z., and Rauen, T. (2005). Individual subunits of the glutamate transporter EAAC1 homotrimer function independently of each other. *Biochemistry* 44, 11913–11923.
- Groeneveld, M., and Slotboom, D.J. (2010). Na⁺:aspartate coupling stoichiometry in the glutamate transporter homologue GlT_{PH}. *Biochemistry* 49, 3511–3513.
- Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008). GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4, 435–447.
- Jensen, M.O., Jogini, V., Borhani, D.W., Leffler, A.E., Dror, R.O., and Shaw, D.E. (2012). Mechanism of voltage gating in potassium channels. *Science* 336, 229–233.
- Kanner, B.I., and Sharon, I. (1978). Active transport of L-glutamate by membrane vesicles isolated from rat brain. *Biochemistry* 17, 3949–3953.
- Kovermann, P., Machtens, J.P., Ewers, D., and Fahlke, C. (2010). A conserved aspartate determines pore properties of anion channels associated with excitatory amino acid transporter 4 (EAAT4). *J. Biol. Chem.* 285, 23676–23686.
- Kutzner, C., Grubmüller, H., de Groot, B.L., and Zachariae, U. (2011). Computational electrophysiology: the molecular dynamics of ion channel permeation and selectivity in atomistic detail. *Biophys. J.* 101, 809–817.
- Lakowicz, J.R. (2006). Principles of fluorescence spectroscopy. In *Principles of Fluorescence Spectroscopy*, P.H. Lakowicz, ed. (New York: Springer), pp. 529–575.
- Machtens, J.P., Kovermann, P., and Fahlke, C. (2011a). Substrate-dependent gating of anion channels associated with excitatory amino acid transporter 4. *J. Biol. Chem.* 286, 23780–23788.
- Melzer, N., Biela, A., and Fahlke, C. (2003). Glutamate modifies ion conduction and voltage-dependent gating of excitatory amino acid transporter-associated anion channels. *J. Biol. Chem.* 278, 50112–50119.
- Mim, C., Balani, P., Rauen, T., and Grewer, C. (2005). The glutamate transporter subtypes EAAT4 and EAATs 1–3 transport glutamate with dramatically different kinetics and voltage dependence but share a common uptake mechanism. *J. Gen. Physiol.* 126, 571–589.
- Otis, T.S., and Kavanaugh, M.P. (2000). Isolation of current components and partial reaction cycles in the glial glutamate transporter EAAT2. *J. Neurosci.* 20, 2749–2757.
- Picaud, S.A., Larsson, H.P., Grant, G.B., Lecar, H., and Werblin, F.S. (1995). Glutamate-gated chloride channel with glutamate-transporter-like properties in cone photoreceptors of the tiger salamander. *J. Neurophysiol.* 74, 1760–1771.
- Reyes, N., Ginter, C., and Boudker, O. (2009). Transport mechanism of a bacterial homologue of glutamate transporters. *Nature* 462, 880–885.
- Ryan, R.M., and Mindell, J.A. (2007). The uncoupled chloride conductance of a bacterial glutamate transporter homolog. *Nat. Struct. Mol. Biol.* 14, 365–371.
- Ryan, R.M., Mitrovic, A.D., and Vandenberg, R.J. (2004). The chloride permeation pathway of a glutamate transporter and its proximity to the glutamate translocation pathway. *J. Biol. Chem.* 279, 20742–20751.
- Ryan, R.M., Kortt, N.C., Sirivanta, T., and Vandenberg, R.J. (2010). The position of an arginine residue influences substrate affinity and K⁺ coupling in the human glutamate transporter, EAAT1. *J. Neurochem.* 114, 565–575.

- Schneider, N., Cordeiro, S., Machtens, J.P., Braams, S., Rauen, T., and Fahlke, C. (2014). Functional properties of the retinal glutamate transporters GLT-1c and EAAT5. *J. Biol. Chem.* 289, 1815–1824.
- Shabaneh, M., Rosental, N., and Kanner, B.I. (2014). Disulfide cross-linking of transport and trimerization domains of a neuronal glutamate transporter restricts the role of the substrate to the gating of the anion conductance. *J. Biol. Chem.* 289, 11175–11182.
- Shrivastava, I.H., Jiang, J., Amara, S.G., and Bahar, I. (2008). Time-resolved mechanism of extracellular gate opening and substrate binding in a glutamate transporter. *J. Biol. Chem.* 283, 28680–28690.
- Torres-Salazar, D., and Fahlke, C. (2007). Neuronal glutamate transporters vary in substrate transport rate but not in unitary anion channel conductance. *J. Biol. Chem.* 282, 34719–34726.
- Vaitheeswaran, S., Rasaiah, J.C., and Hummer, G. (2004). Electric field and temperature effects on water in the narrow nonpolar pores of carbon nanotubes. *J. Chem. Phys.* 121, 7955–7965.
- Vázquez-Ibar, J.L., Guan, L., Weinglass, A.B., Verner, G., Gordillo, R., and Kaback, H.R. (2004). Sugar recognition by the lactose permease of *Escherichia coli*. *J. Biol. Chem.* 279, 49214–49221.
- Verdon, G., and Boudker, O. (2012). Crystal structure of an asymmetric trimer of a bacterial glutamate transporter homolog. *Nat. Struct. Mol. Biol.* 19, 355–357.
- Verdon, G., Oh, S., Serio, R.N., and Boudker, O. (2014). Coupled ion binding and structural transitions along the transport cycle of glutamate transporters. *eLife* 3, e02283.
- Wadiche, J.I., and Kavanaugh, M.P. (1998). Macroscopic and microscopic properties of a cloned glutamate transporter/chloride channel. *J. Neurosci.* 18, 7650–7661.
- Wersinger, E., Schwab, Y., Sahel, J.A., Rendon, A., Pow, D.V., Picaud, S., and Roux, M.J. (2006). The glutamate transporter EAAT5 works as a presynaptic receptor in mouse rod bipolar cells. *J. Physiol.* 577, 221–234.
- Winter, N., Kovermann, P., and Fahlke, C. (2012). A point mutation associated with episodic ataxia 6 increases glutamate transporter anion currents. *Brain* 135, 3416–3425.
- Yernool, D., Boudker, O., Jin, Y., and Gouaux, E. (2004). Structure of a glutamate transporter homologue from *Pyrococcus horikoshii*. *Nature* 431, 811–818.

Enhancer Evolution across 20 Mammalian Species

Graphical Abstract



Authors

Diego Villar, Camille Berthelot, ..., Paul Flicek, Duncan T. Odom

Correspondence

flicek@ebi.ac.uk (P.F.),
duncan.odom@cruk.cam.ac.uk (D.T.O.)

In Brief

Comparative functional genomic analysis in 20 mammalian species reveals distinct features for the evolution of enhancers, in comparison to those of promoters, across 180 million years.

Highlights

- Rapid enhancer and slow promoter evolution across genomes of 20 mammalian species
- Enhancers are rarely conserved across these mammals
- Recently evolved enhancers dominate mammalian regulatory landscapes
- Unbiased mapping links candidate enhancers with lineage-specific positive selection



Enhancer Evolution across 20 Mammalian Species

Diego Villar,^{1,11} Camille Berthelot,^{2,11} Sarah Aldridge,^{1,12} Tim F. Rayner,¹ Margus Lukk,¹ Miguel Pignatelli,² Thomas J. Park,³ Robert Deaville,⁴ Jonathan T. Erichsen,⁵ Anna J. Jasinska,⁶ James M.A. Turner,⁷ Mads F. Bertelsen,⁸ Elizabeth P. Murchison,⁹ Paul Flicek,^{2,10,*} and Duncan T. Odom^{1,10,*}

¹University of Cambridge, Cancer Research UK Cambridge Institute, Robinson Way, Cambridge, CB2 0RE, UK

²European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK

³Department of Biological Sciences, University of Illinois at Chicago (UIC), 845 West Taylor Street, Chicago, IL 60607, USA

⁴UK Cetacean Strandings Investigation Programme (CSIP) and Institute of Zoology, Zoological Society of London, Outer Circle, Regent's Park, London NW1 4RY, UK

⁵School of Optometry and Vision Sciences, Cardiff University, Maindy Road, Cardiff CF24 4HQ, UK

⁶UCLA Center for Neurobehavioral Genetics, 695 Charles E. Young Drive South, Los Angeles, CA 90095, USA

⁷Division of Stem Cell Biology and Developmental Genetics, MRC National Institute for Medical Research, Mill Hill, London NW7 1AA, UK

⁸Center for Zoo and Wild Animal Health, Copenhagen Zoo, Roskildevej 38, DK-2000 Frederiksberg, Denmark

⁹Department of Veterinary Medicine, University of Cambridge, Madingley Road, Cambridge CB3 0ES, UK

¹⁰Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK

¹¹Co-first author

¹²Present address: AstraZeneca, Personalised Healthcare and Biomarkers, Molecular Diagnostics, Darwin Building 310, Milton Rd, Cambridge CB4 0WG, UK

*Correspondence: flicek@ebi.ac.uk (P.F.), duncan.odom@cruk.cam.ac.uk (D.T.O.)

<http://dx.doi.org/10.1016/j.cell.2015.01.006>

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

SUMMARY

The mammalian radiation has corresponded with rapid changes in noncoding regions of the genome, but we lack a comprehensive understanding of regulatory evolution in mammals. Here, we track the evolution of promoters and enhancers active in liver across 20 mammalian species from six diverse orders by profiling genomic enrichment of H3K27 acetylation and H3K4 trimethylation. We report that rapid evolution of enhancers is a universal feature of mammalian genomes. Most of the recently evolved enhancers arise from ancestral DNA exaptation, rather than lineage-specific expansions of repeat elements. In contrast, almost all liver promoters are partially or fully conserved across these species. Our data further reveal that recently evolved enhancers can be associated with genes under positive selection, demonstrating the power of this approach for annotating regulatory adaptations in genomic sequences. These results provide important insight into the functional genetics underpinning mammalian regulatory evolution.

INTRODUCTION

Most mammalian genes are controlled by collections of enhancer regions, often located tens to hundreds of kilobases away from transcription start sites. Recent studies comparing key selected mammals (Cotney et al., 2013; Xiao et al., 2012) have indicated that enhancers may change rapidly during evolu-

tion (Degner et al., 2012; Shibata et al., 2012), particularly when compared with evolutionarily stable gene expression patterns (Brawand et al., 2011; Chan et al., 2009; Merkin et al., 2012). Given that most phenotypic differences are hypothesized to largely result from regulatory differences between mammals, it is of profound importance to understand the mechanisms driving enhancer evolution (Villar et al., 2014; Wray, 2007).

Both conserved and recently evolved enhancer sequences have been shown to have important phenotypic consequences. Highly conserved enhancer sequences can regulate fundamental processes, such as embryonic development, and this property has been used to screen for functional regulatory elements (Pennacchio et al., 2006). However, sequence-level changes in enhancer elements can also underlie evolutionary differences between species (Hare et al., 2008; Ludwig et al., 2005), as has now been demonstrated across many organisms (Arnold et al., 2014; Cotney et al., 2013; Degner et al., 2012; McLean et al., 2011; Shibata et al., 2012).

Approaches comparing vertebrate genome sequences, such as those employing 29 mammals, have revealed regulatory regions under sequence constraint (Lindblad-Toh et al., 2011). However, this approach is limited in resolving tissue-specific deployment or regulatory activity directed by small sequence changes, particularly as may be predicted for rapidly evolving enhancer regions (however, see Pollard et al., 2006; Prabhakar et al., 2006). Comparative analysis of mammalian genomes can indicate protein sequence adaptations in particular species or lineages, and infer which coding regions are under positive selection. In contrast, complementary experimental efforts are currently lacking to functionally annotate the many recently sequenced mammalian genomes.

Experimental tools can now empirically identify regulatorily active DNA across entire mammalian genomes. Enhancers can

be identified by mapping regions enriched for acetylated lysine 27 on histone H3 (H3K27ac) via chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) (Creyghton et al., 2010). Similarly, active gene promoters can be identified as containing both H3K27ac and trimethylated lysine 4 of histone H3 (H3K4me3), which marks sites of transcription initiation (Cain et al., 2011; Santos-Rosa et al., 2002). The usefulness of this approach to map regulatory activity genome-wide has been recently underscored by analysis of H3K27ac dynamics across organ development in mouse (Nord et al., 2013). This study found that most H3K27ac developmental variation occurs distally to transcription start sites and within predicted enhancer elements, most of which could be validated experimentally.

Over 20 sequenced mammalian genomes have been integrated into inter-species alignments within Ensembl (Flicek et al., 2014). Exploiting this computational infrastructure (and related resources in *Drosophila*; Kim et al., 2009), recent studies have dissected how transcription factor (TF) binding has evolved (He et al., 2011; Paris et al., 2013; Schmidt et al., 2010; Stefflova et al., 2013). In addition, enhancer and promoter evolution have been investigated using sets of mammals, where H3K27ac levels have been characterized across tissues and developmental states as a proxy for enhancer function and developmental or tissue-specific gene expression (Cotney et al., 2013; Nord et al., 2013; Xiao et al., 2012).

Here, we report the results of empirically mapping promoter and enhancer evolution across 20 mammals chosen to span the breadth and depth of the class Mammalia, including previously uncharacterized species such as cetaceans and naked mole rat. Our analyses have revealed the tempo and mechanisms underlying enhancer evolution across over 180 million years of mammalian radiation.

RESULTS

Profiling Promoter and Enhancer Regulatory Evolution in Mammalian Liver

We mapped the active promoter and enhancer elements in liver as a representative adult somatic tissue from 20 species of mammals (Figure 1). Study species were selected using three criteria: (1) to capture a substantial fraction of the mammalian phylogenetic tree, (2) to profile the major placental orders in a combination of intra- (6–40 Ma) and inter-lineage (100–180 Ma) evolutionary distances, and (3) to extend our understanding of regulatory evolution to previously uncharacterized mammals whose phenotypes are highly divergent, such as cetaceans, naked mole rat, and Tasmanian devil. Liver from almost all study species was profiled in biological replicates from two or more individuals, except for Sei Whale (*Balaenoptera borealis*), where only one individual's tissue was available; and for dolphin, for which we combined data from two closely related dolphin species (*Delphinus delphis* and *Lagenorhynchus albirostris*) where a single individual from each species was profiled (Tables S1 and S2, Experimental Procedures).

We quantified using ChIP-seq the genome-wide occurrence of two key histone marks widely used to profile promoters and enhancers: H3K4me3 and H3K27ac (Figure 1) (Creyghton

et al., 2010; Santos-Rosa et al., 2002). We identified regions enriched for these histone marks within each mammalian liver genome using only biologically reproducible peaks present in two or more replicates (Figure S1, Experimental Procedures).

A total of 30–45,000 regions per species were enriched in liver, and these separated into H3K27ac, H3K4me3&H3K27ac, and H3K4me3-marked elements (Figures 1C and S1). Our analyses were robust to variability in the genome assembly quality and sample preparation (Experimental Procedures and Figure S2). We confirmed that H3K4me3 often co-occupied the genome with H3K27ac (Heintzman et al., 2009; Zhu et al., 2013), and that most H3K4me3-positive regions occur at transcriptional start sites (Cain et al., 2011; Santos-Rosa et al., 2002), regardless of their H3K27ac enrichment (see Experimental Procedures). In contrast, regions enriched for H3K27ac often were not enriched for H3K4me3, and these often located far from transcriptional start sites (Figure S2).

The regions we identify as enhancers strongly enrich for regulatory activity in liver, consistent with numerous prior studies (Cotney et al., 2013; Creyghton et al., 2010; Nord et al., 2013; Zhu et al., 2013). For over 400 of our human liver enhancers (typically 2 kb in length), the transgenic activities of overlapping 145 bp segments were assayed in liver cancer cells (Kheradpour et al., 2013) (Figure S2). Although each human liver enhancer was on average represented by only a single small sequence element, capturing less than 10% of the enhancer length, over 65% showed activity in transgenic assays in a cancer cell line. Furthermore, over 90% of the enhancers not active in transgenic assays were nevertheless bound in human liver by at least one liver-specific TF (Ballester et al., 2014). In sum, this analysis suggests a sizable majority of our empirically determined enhancers are regulatorily active.

Our data newly demonstrates that the known interplay of H3K4me3 and H3K27ac creates a genomic regulatory landscape that is a uniform feature across mammals (and likely across eumetazoans; Schwaiger et al., 2014). In adult liver, a typical mammalian genome contains on average 12,500 H3K4me3 locations (representing active promoter elements) and 22,500 H3K27ac-enriched regions (representing active enhancers).

Enhancer Evolution Is Appreciably More Rapid Than Proximal Promoter Evolution

We used our genome-wide mapping data in livers from 20 mammals to obtain an empirical and quantitative understanding of evolutionary stability of promoters and enhancers (Figure 2 and Figure S3).

Most non-coding regions in the human genome cannot be mapped across 20 mammals, in large part because the genome structure and regulatory content of complex eukaryotes evolve rapidly (Lynch et al., 2011). We defined the maximum detectable conservation of activity as the number of species in which the DNA could be aligned (Figure 2A). For example, if enhancer activity is highly conserved, then this activity would be detected in all species where the underlying DNA was alignable. In contrast, low conservation would be characterized by the underlying DNA remaining alignable across many species, but without sharing of enhancer activity. Such low conservation could be a

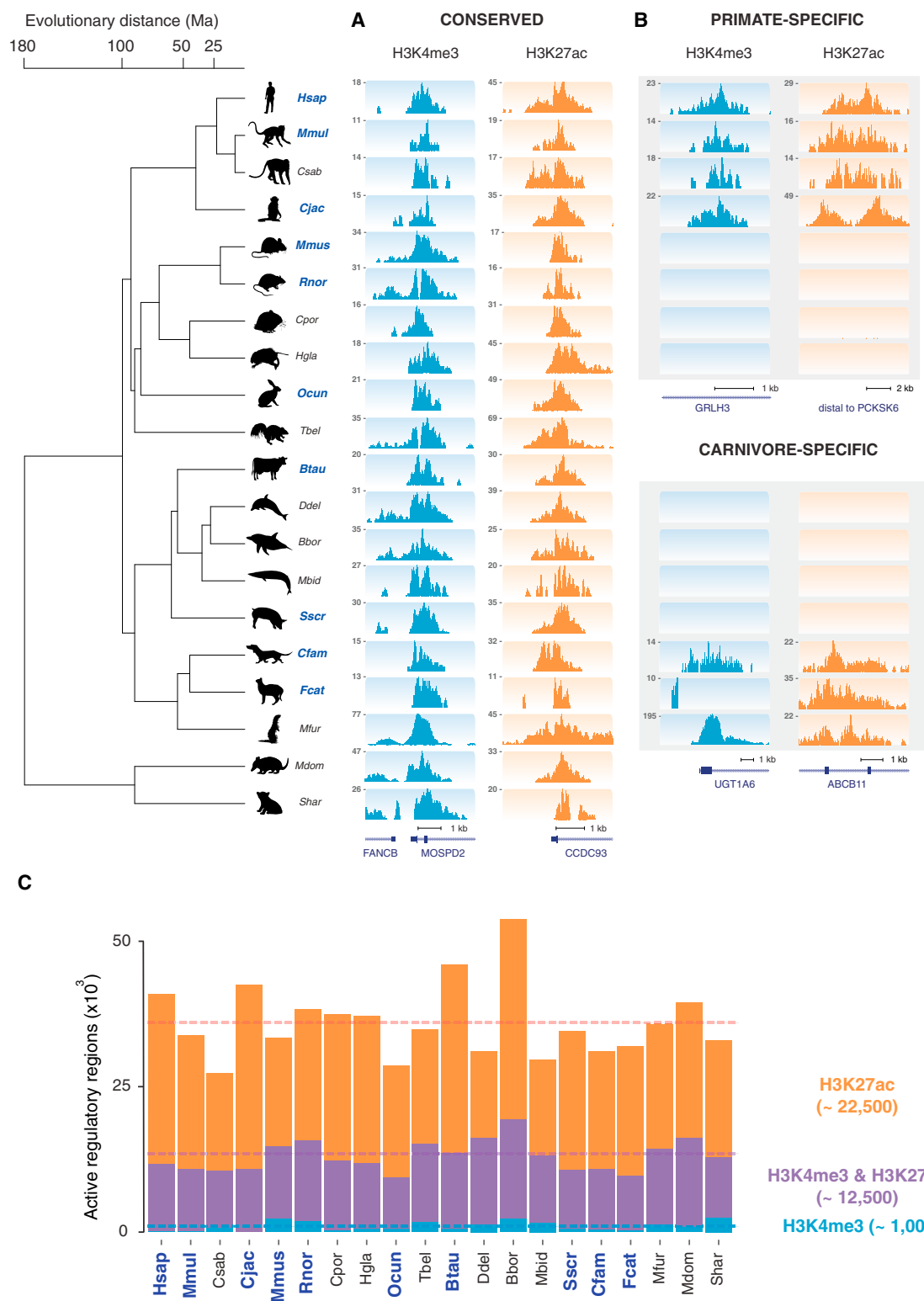


Figure 1. In Vivo Regulatory Activity Assessed in Livers from 20 Mammals
(A and B) Phylogenetic relationships and species divergences are represented by an evolutionary tree, which includes 18 placental species (in four orders) and 2 marsupial species (in two orders). In liver isolated from each species, enhancer activity was globally mapped by identifying genomic regions enriched for acetylation of H3K27 (H3K27ac), and transcription initiation was mapped by identifying genomic regions enriched for tri-methylation of H3K4 (H3K4me3). Shown (legend continued on next page)

signature of rapid functional evolution or, alternatively, functional neutrality.

Collectively, the DNA sequences used as promoters and the DNA sequences used as enhancers in liver show only slight differences in their alignability across the study species (Figure 2B). This alignability shows a marked increase at approximately 11–13 species, reflecting the contribution to the multiple alignments of the ten highest-quality genomes (Experimental Procedures).

The conservation of active liver promoters tracked remarkably closely with the alignability of the underlying DNA, indicating evolutionarily stable promoter activity (Figure 2C, upper left triangle). In other words, the transcription initiation sites driving gene expression in liver are highly conserved.

We performed a similar analysis for enhancers. Our data reveal that rapid enhancer evolution, often involving exaptation of ancestral DNA, is active and widespread across all the mammalian clades in our study (Figure 2D, orange, and Figure S3), as has been reported in primates (Cotney et al., 2013). Furthermore, the ten highest-quality placental genome sequences contained thousands of cross-alignable regions where enhancer activity was shared in many, but not all, species. These regions are liver enhancers that were likely present in the common placental ancestor and have partially degraded along some lineages. In contrast to promoter sites, enhancer locations evolve rapidly, and comparatively few are deeply conserved (see below). Control analyses show that while promoter conservation may be under-estimated, this is not the case for enhancers (Figure S3).

We asked whether the conservation of liver promoters and enhancers is associated with underlying sequence features (e.g., TF binding sequences, %GC content, sequence constraint), experimental features (reproducibility, occupancy level/intensity, length), or some combination (Figure 3). The best predictor of conservation in promoter regions is the reproducibility and strength of enrichment of H3K4me3 and H3K27ac, with the length of the histone-modified domain and GC content as separate, modest contributors. Thus, experimental features are stronger indicators of the conservation of regulatory activity, and underlying sequence features contribute less to promoter stability. In contrast, the presence of TF binding sites can explain a modest fraction of the conservation of enhancer activity. Nevertheless, as with promoters, the enrichment reproducibility and intensity of signal is the primary predictor of conservation. Collectively, no combination of sequence- and experimental-based features could potentially explain more than a third of the variance in conservation of regulatory activity.

Overall, our data reveal that promoter activity in a representative somatic tissue is highly constrained across mammalian space. In contrast, enhancer evolution is rapid and widespread. Neither enhancer nor promoter activity conservation can be explained purely by underlying sequence elements.

Quantifying the Divergence Rates of Enhancers, Promoters, and TF Binding in a Cross-Section of Mammals

The divergence rate of sequence-specific transcription factor binding (Stefflova et al., 2013) and the extent of regulatory evolution (Cotney et al., 2013; Shibata et al., 2012; Xiao et al., 2012) has been estimated using matched experiments from the same tissues in subsets of typically three to five mammals within a single order. We took a similar approach to calculate how rapidly enhancers and promoters active in liver evolve across 20 mammals.

We first identified, by pairwise analysis of all 20 species, whether regions called as enhancers and promoters were present in the same location between two mammalian genomes (Experimental Procedures, Figure S4). Because this analysis does not use human as the primary reference genome, we could generate multiple independent estimates of how evolutionarily stable enhancers and promoters were for comparable divergence distances. Further, divergence rates could be estimated for evolutionary distances not available from a human-centric analysis. For instance, our data provided multiple comparisons of species separated by 40 to 100 Ma using mouse, cow, or dog as reference that could not be obtained using a human-centric approach (Figure 1).

Inter-species conservation of promoters and enhancers could be plausibly described as a function of time-of-divergence by fitting an exponential decay curve (Experimental Procedures). In liver, promoters diverged at a slower rate than did either enhancers or TF bound regions (Figure 4 and Figure S4). Interestingly, promoters' half-lives are comparable to protein-coding genes' half-lives, at over a billion years (Rands et al., 2014). The higher stability of promoters versus enhancers could be due in part to the intimate functional connection promoters have with the first exon of protein coding genes, which are highly stable features of vertebrate genomes (Lindblad-Toh et al., 2011). Our results are consistent with a model where the increased size and sequence heterogeneity of regions with promoter or enhancer activity could buffer evolutionary changes more robustly than can site-specific TF binding alone (Cotney et al., 2013; Shibata et al., 2012; Xiao et al., 2012).

Highly Conserved Regulatory Regions Are Largely Proximal Promoters

Our mapping of liver enhancer and promoter evolution using mammals spanning both intra-order (6–40 Ma) and inter-order (80–180 Ma) divergence times permits the dissection of conserved (and recently evolved, see below) regulatory regions.

We first quantified how many regions showed strong conservation of activity by defining regions as *highly conserved* if regulatory activity was present in (at a minimum) all ten of the highest-quality placental genomes (Figure 5A). A total of 2,151 genomic regions appeared highly conserved by these criteria, representing 5% of all human regions active in liver. The

are examples of regulatory regions active: (A) across all 20 species (MOSPD2 and CCDC93 loci), and (B) active only in primates (GRLH3 and PCKSK8, top) or active only in carnivores (UGT1A6 and ABCB11, bottom). For order-specific regulatory regions, data from some species are not shown for conciseness. (C) In liver, a typical mammalian genome contains ~22,500 enhancers enriched for only H3K27ac; ~12,500 promoters enriched for both H3K27ac and H3K4me3 and ~1,000 containing only H3K4me3. Highest quality genomes incorporated into the EPO multiple alignment are labeled in blue (Experimental Procedures). See also Figures S1 and S2 and Tables S1 and S2.

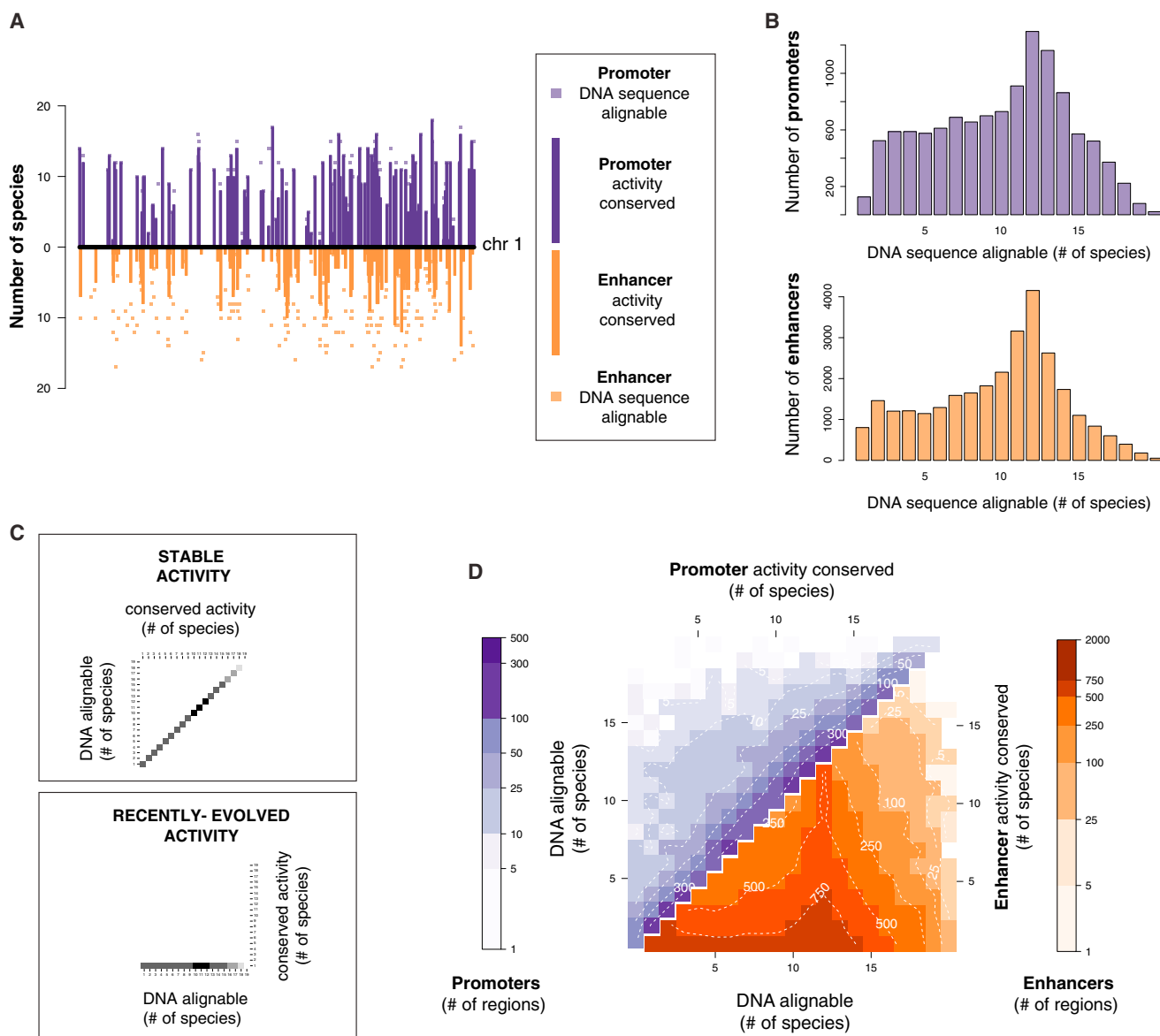


Figure 2. Enhancers Evolve Rapidly; Promoters Are Highly Conserved

(A) For a representative 10 MB region on human chromosome 1, the bar chart on the y axis represents the number of species in which enhancer and promoter elements were active (promoters: top, purple; enhancers: bottom, orange). Squares indicate the number of species where the sequence underlying the active promoter or enhancer was alignable.

(B) The DNA sequences underlying proximal promoters and the DNA sequences underlying enhancers can be aligned to similar numbers of species, suggesting that differences in apparent conservation of activity are not due to differences in alignability.

(C) Schematic diagram showing how the conservation of regulatory activity versus DNA alignability across 20 species of mammals can reveal (top) where DNA function and DNA sequence orthology closely correspond, indicating ancestral activity, and (bottom) where pre-existing DNA sequences have been exapted within specific lineages or species, indicating recently evolved activity.

(D) Our data revealed that if the DNA underlying a human-identified proximal promoter region (purple) can be aligned with an orthologous sequence in another species, then promoter activity is very often present as well (heatmap enrichment concentrated on the diagonal of the plot). In contrast, most enhancer regions (orange) are rapidly evolving within older DNA sequences, reflected in increased heatmap enrichment toward the lower x axis. Color scales and dashed contour lines indicate absolute numbers of active promoter or enhancer regions (logarithmic scale).

See also Figure S3.

existence of over 2,000 highly conserved regions is greater than expected by chance (p value $< 1 \times 10^{-4}$, random permutation test, [Experimental Procedures](#)).

Highly conserved regions were classified as promoters or enhancers based on their consensus histone mark enrichment across all 20 mammals ([Experimental Procedures](#)). Of these

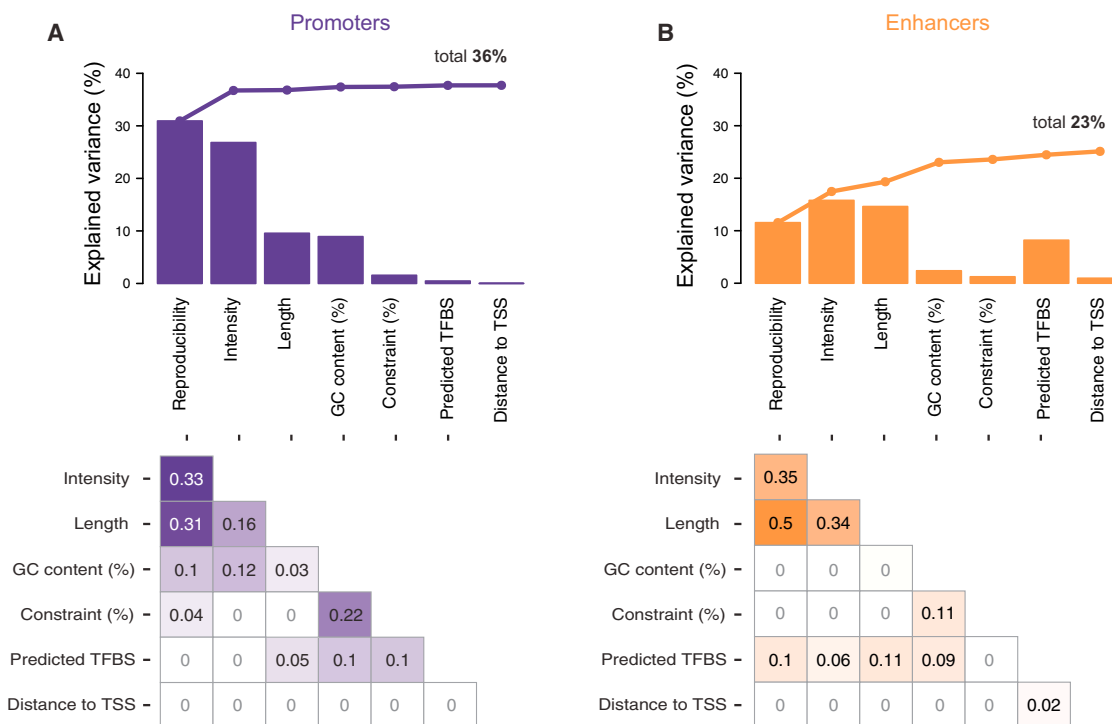


Figure 3. Features Contributing to Conservation of Promoter and Enhancer Activity Identified in Human Liver

(A) For all human proximal promoters active in liver, the depth of conservation was correlated with experimental features (reproducibility, peak intensity, peak length, distance to nearest transcription start site) as well as underlying genomic features (GC content, sequence constraint, TF binding sites). Each feature in isolation explained a significant fraction of the variance in conservation of promoter activity (e.g., peak length explained 10%). The fraction explained by the features in combination, when added left to right using multiple regression analysis, are plotted as a line above, in sum totaling 36%. The increases in explained variance with the addition of each feature are attenuated due to strong inter-correlation of features, quantified in the bottom panel as R^2 values between features (Experimental Procedures).

(B) The same analysis was performed for human liver enhancers, where experimental and genomic features together explained a more modest fraction (23%) of the conservation of enhancer activity in other species.

2,151 highly conserved regulatory regions, 1,871 elements (87%) were enriched for both H3K27ac and H3K4me3, consistent with acting as promoters (Santos-Rosa et al., 2002). The vast majority of highly conserved promoters occupied the transcription start sites of genes (Figure 5B). On the other hand, a subset of 279 regions showed enrichment only for H3K27ac occupancy, consistent with acting as enhancers (Creyghton et al., 2010). Most highly conserved enhancers were tens to hundreds of kilobases away from the nearest gene (Figure 5B). The single region uniformly enriched across placentals for only H3K4me3 is not shown.

In human liver, there are 11,838 promoter regions enriched for both H3K27ac and H3K4me3, and 28,963 enhancer regions containing only H3K27ac. Although nearly three times as common as promoters, the activity of only 1% of these enhancers is highly conserved. In contrast, the activity of 16% of promoters is highly conserved (Figure 5A).

Three independent lines of evidence support the functionality of the sequences we identify as highly conserved regulatory regions in liver. First, all show enhanced sequence constraint (Figure 5C). Second, genes near highly conserved enhancers are strongly enriched for liver-specific functions, and genes near conserved proximal promoters are enriched for house-

keeping functions (Figure S5, Tables S3 and S6) (Forrest et al., 2014). Third, highly conserved enhancers are enriched for TF binding motifs for liver-specific regulators such as CEBPA and PBX1, whereas highly conserved proximal promoters appear dominated by transcriptional initiation regulatory sequences (Figure S5, Table S7).

In sum, in adult mammals comparatively few enhancers are evolutionarily stable. In contrast, a substantial fraction of the proximal promoters found in human liver appear to be highly conserved across mammals.

Recently Evolved Regulatory Activity Is Pervasive in Mammals

Even for proximal promoters, the number of highly conserved regulatory elements active in liver is a small fraction of the total number experimentally identified in any single species (Figure 5 and Table S4). We sought to identify and analyze the molecular features of more recently evolved regulatory regions.

From each placental order, we selected a representative species (human, mouse, cow, dog) and then identified a set of newly evolved or, more formally, apomorphic active promoters and enhancers in liver (Figure 6 and Figure S7). For each of these four species, we started with all active regions and then removed

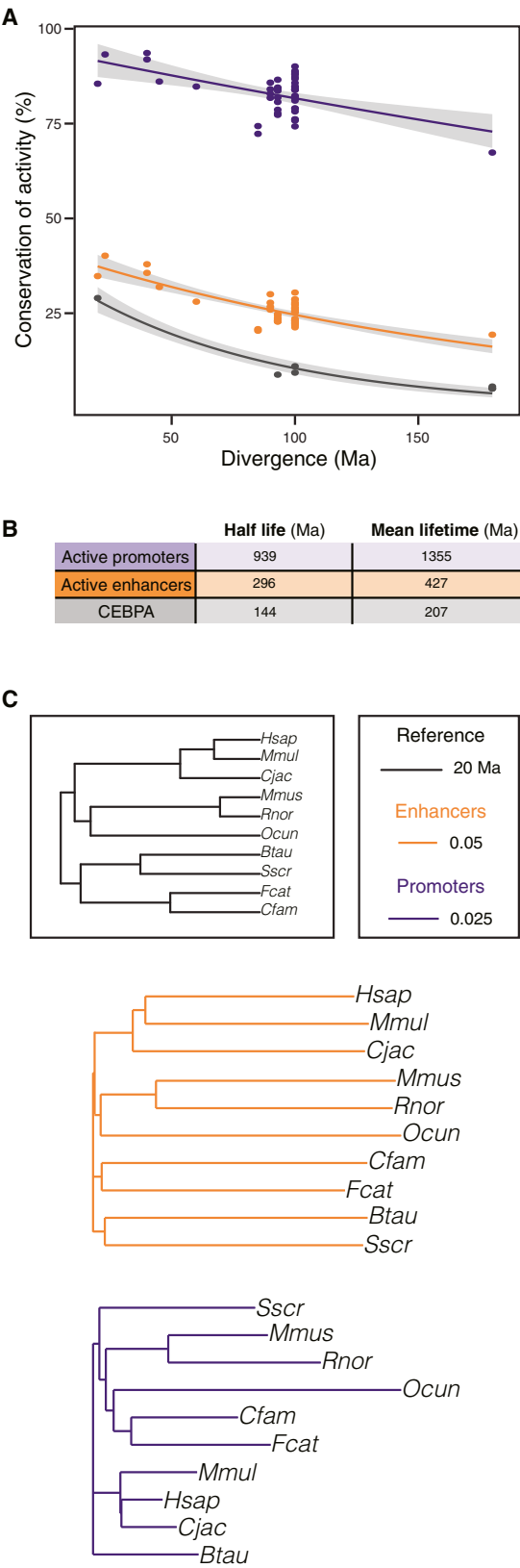


Figure 4. Empirically Determined Rates of Promoter, Enhancer, and TF Binding Divergence in Liver across 180 Million Years of Mammalian Evolution

(A) For promoters (purple), enhancers (orange), and TF binding sites (CEBPA, black), the fraction of ChIP-seq peaks present at the orthologous location between pairs of mammals are shown as a function of evolutionary distance. Solid lines represent an exponential decay fit, surrounded by gray shading of a 95% confidence interval (Experimental Procedures). For liver promoters and enhancers, we used data from the ten highest-quality placental genomes, while CEBPA data have been previously reported (Schmidt et al., 2010). (B) Comparative half-lives and mean-lifetimes (in million years) for active promoters, enhancers and CEBPA transcription factor binding locations, as calculated from the exponential decay fits in (A). (C) Neighbor-joining phylogenetic trees based on pairwise conservation levels of enhancer and promoter activity, as measured in (A). Enhancer evolution (orange) recapitulates the known relationships among the studied mammals (black). The low divergence of promoter activity is insufficient to resolve the phylogenetic groups (purple). See also Figure S4.

those that showed any activity within alignable regions in any other study species (see Experimental Procedures). We found that a typical mammalian liver deploys between 1,000 to 2,000 promoters and 10,000 enhancers not found in any other study species; we henceforth refer to these enhancers and promoters as *recently evolved*.

These numbers are comparable to the extent of enhancer gains previously reported in inter-primate comparisons (Cotney et al., 2013; Shibata et al., 2012) and the extent of promoter evolution estimated from mouse-human comparisons (Forrest et al., 2014; Frith et al., 2006). Especially for enhancers, recently evolved regions are 10–20 times more abundant than those conserved across placentals or shared across multiple species in a particular lineage (Table S4). Both highly conserved and recently evolved regulatory regions active in liver are associated with increased expression of neighboring genes (Figure S6).

Exaptation Drives Recently Evolved Enhancer, but Not Promoter, Activity

Using these tens of thousands of apomorphic regulatory regions, we tested whether functional exaptation of ancestral DNA, recently reported for human-specific enhancers active in embryonic limb (Cotney et al., 2013), is a prevalent mechanism in mammalian genome evolution.

We first asked whether recently evolved proximal promoters are primarily found in ancestral DNA sequences older than 100 Ma (Figure 6A, Figure S7). To our surprise, we discovered that across four orders of mammals, the recent evolution of promoters occurred within evolutionarily younger DNA segments (i.e., not shared with other study species) about three to four times as often as occurred by exaptation of ancestral DNA. For instance in mouse, 1,400 recently evolved promoters occurred in DNA sequences present only in this species (i.e., not shared even with rat); in contrast, only 260 recently evolved promoters were found in ancestral DNA.

Within the ancestral DNA commandeered into new promoters, and regardless of species interrogated, diverse ERV repeat elements are over-represented, consistent with previous reports that ERVs are pre-primed to transcriptional initiation (Fort et al., 2014).

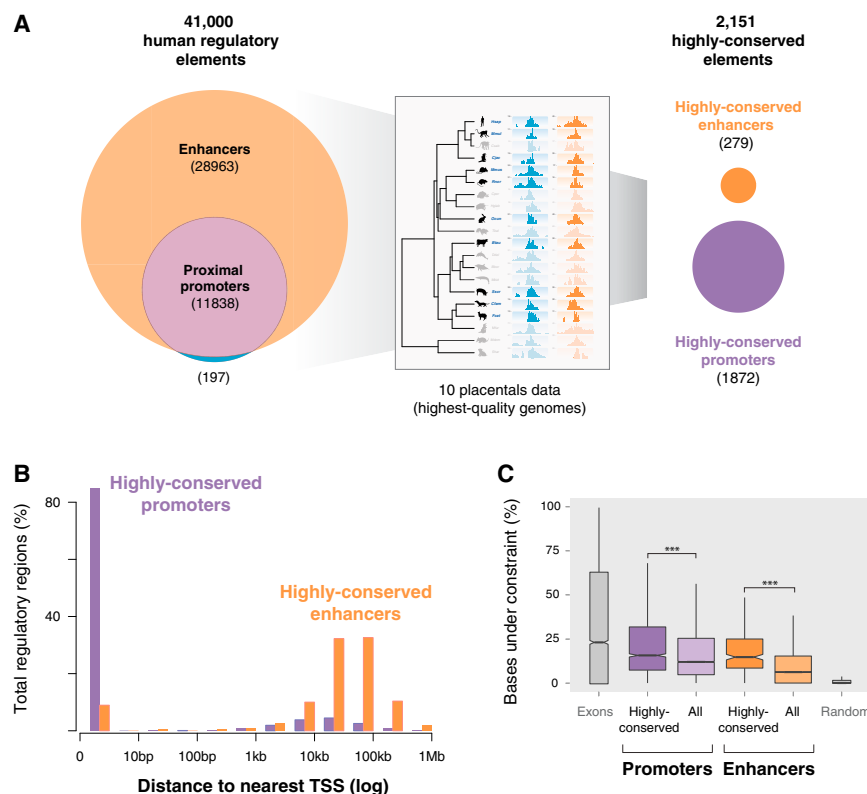


Figure 5. Most Highly Conserved Liver Regulatory Regions Are Proximal Promoters

(A) The ~41,000 regulatorily active regions in human liver are shown on the left panel (enhancers: orange; promoters: purple). The regulatory elements with conserved activity in the ten placental species with highest quality genomes (boxed inset) were determined by cross-species comparison (Experimental Procedures), identifying approximately 300 enhancers and 1,800 promoters (labeled as highly conserved, right panel). (B) Almost all highly conserved promoter regions (purple) are located at transcription start sites as expected, whereas conserved enhancer regions (orange) are typically tens to hundreds of kilobases from the nearest gene.

(C) Regions of highly conserved enhancer and promoter activity show a corresponding, but modest, increase in selective constraint in their underlying DNA sequence. The distribution of the fraction of bases under constraint in each region within each category is shown as a box-plot, with human exons and randomly selected regions shown for comparison (Experimental Procedures).*** indicates p value $< 2 \times 10^{-16}$, Wilcoxon test.

See also Figures S5 and S6 and Tables S3, S6, and S7.

In contrast, the vast majority of enhancers in liver are recently evolved (Table S4)—as well as far more likely to exapt ancestral DNA (Figure 6B). Of the typically 10,000 recently evolved enhancers in a given species, 52%–77% contained sequences of ancestral DNA over 100 Ma old. The remaining recently evolved enhancers were found in younger DNA, and enriched for mobile repetitive element families, including LTRs in all lineages and lineage-specific SINEs and DNA transposons exclusive to primates, carnivores, or ungulates (Figure 6B).

In a typical mammalian species, the 1,000 to 2,000 recently evolved liver promoters occur predominantly in younger DNA typically less than 40 Ma old, whereas the 10,000 recently evolved enhancers are formed predominantly by exaptation of ancestral DNA. Only a minority of recently evolved enhancers and promoters appear driven by repeat element expansions (Figure 6, Figure S7). Across our study's 20 mammals, exaptation of ancestral DNA generates more of the recently evolved regulatory genome than do repeat-driven expansions.

Functional Annotation of Genes under Positive Selection

Comparing genome sequences can suggest which genes drive phenotypic adaptations by using inference of regions under positive selection and by analyzing amino acid substitution patterns in proteins (Nielsen et al., 2007). Both approaches primarily employ coding-sequence alignments and thus provide limited insight into regulatory adaptations. We therefore asked whether genes under positive selection are associated with apomorphic enhancers, perhaps evolving synergistically (Shibata et al., 2012).

We compared recently evolved enhancers and positively selected genes in two newly sequenced species: (1) naked mole rat, a cancer-resistant rodent (Kim et al., 2011); and (2) dolphin, a marine mammal metabolically adapted to an aquatic environment (Sun et al., 2013). In both species, we found that recently evolved enhancers are over-represented near positively selected genes (Experimental Procedures) (p values = 0.022 [naked mole rat] and 0.023 [dolphin], hypergeometric test. See Table S5).

Illustrative examples are shown in Figure 7. First, a recently evolved enhancer in naked mole rat is shown upstream of the *thymopoietin* gene (*TMPO*), identified previously as positively selected (Kim et al., 2011). The orthologous *TMPO* regions in human, mouse, cow, and dog show no enhancer activity, though a number of partially conserved enhancers are present nearby (Figure 7A). Second, the genomic region around the *TRIP12* gene, under positive selection in dolphin (Sun et al., 2013), contains a recently evolved dolphin enhancer not active in human, mouse, dog, and cow. Moreover, this regulatory element appears to be the main enhancer in this region (Figure 7B).

In sum, recently evolved active regions identified in this study, and in particular rapidly evolving enhancers, can functionally annotate lineage-specific adaptations.

DISCUSSION

We experimentally dissected the evolution of regulatory regions in mammalian liver by mapping the genome-wide landscape of

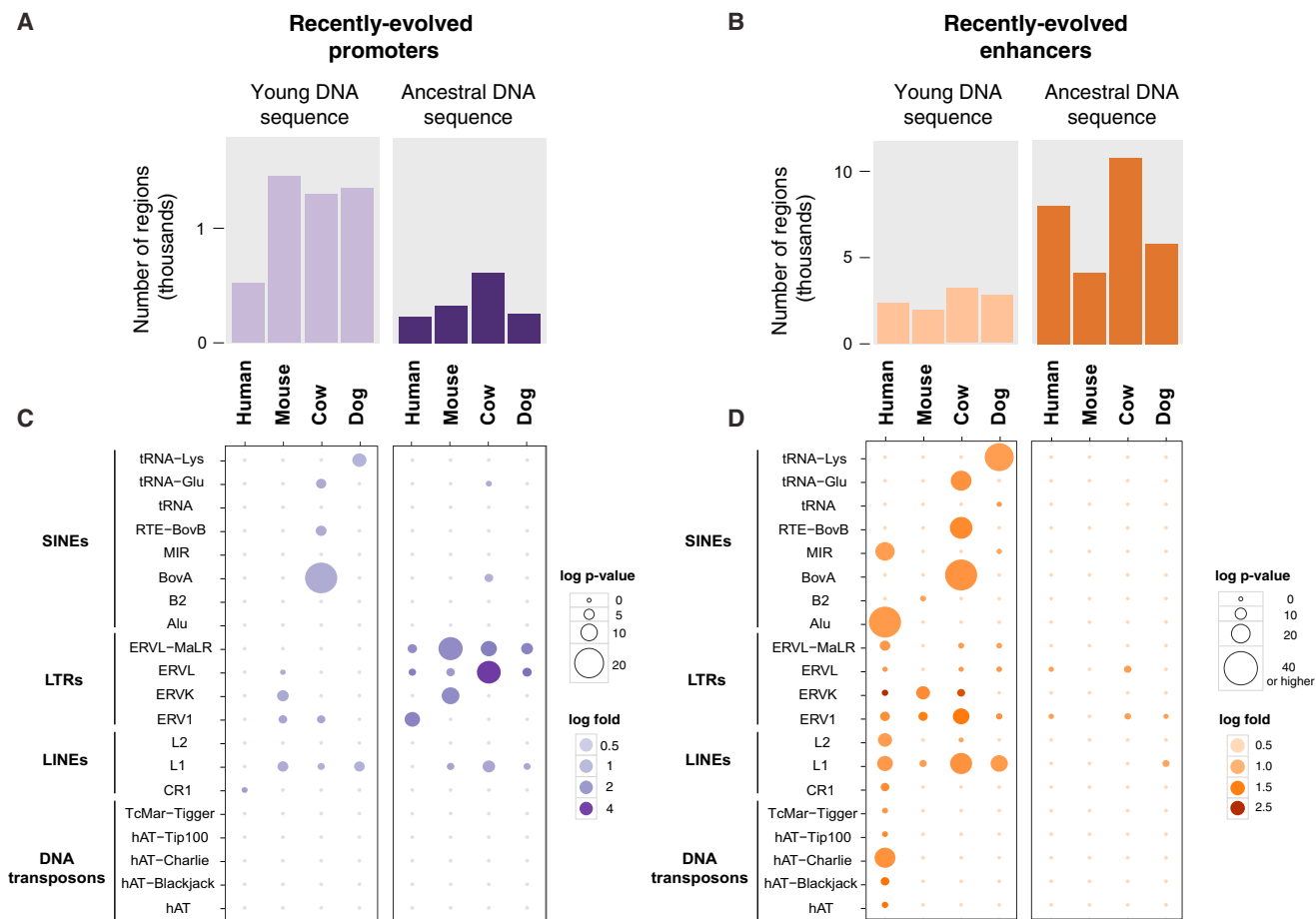


Figure 6. Recently Evolved Promoters Are Largely Derived from Young DNA, While Recently Evolved Enhancers Are Mostly Exapted from Ancestral DNA Sequences

Regions with recently evolved promoter and enhancer activity in liver were identified in a representative species for each placental order (primate:human, rodent:mouse, ungulate:cow, and carnivore:dog). These regions were categorised into those falling in (1) young DNA sequences (0–40 Ma) or (2) ancestral DNA sequences (>100 Ma).

(A) Typically three times as many recently evolved active promoters reside in young DNA as are found in ancestral DNA sequences present across placental mammals.

(B) Conversely, typically twice as many recently evolved enhancers are exapted from evolutionarily ancestral DNA as are found in young DNA.

(C and D) Repeat classes and families enriched in recently evolved promoters and enhancers were identified using a binomial test (see [Experimental Procedures](#)). Plots show enrichments for each repeat family (y axis) and each species (x axis). Circle sizes represent the statistical significance of enrichment, and color shades denote the fold change of the enrichment (both in logarithmic scale).

See also [Figures S6 and S7](#) and [Tables S3, S4, S6, and S7](#).

active promoters and enhancers from 20 diverse species. The evolutionary distances spanning four distinct orders within class *Mammalia* enabled rigorous analysis of the mechanisms underlying regulatory evolution. The combination of rapid enhancer and slower promoter evolution appears to be a fundamental property of the mammalian regulatory genome, shared by species separated by up to 180 million years. A sizable number of the 10,000–15,000 active promoters are functionally shared across most mammals, and are associated with ubiquitous cellular functions; highly conserved enhancers are much less common, and are found near liver-specific genes. Remarkably, almost half of 20,000–25,000 active enhancers in each species have rapidly evolved in a lineage- or species-specific manner. Our genome-wide mapping of enhancers in previously unchar-

acterized species has enabled us to identify regulatory regions near genes under positive selection that may help drive phenotypic adaptations.

A Global Overview of Enhancer and Promoter Evolution in Mammals

We used a powerful and unbiased strategy to confirm, extend, and explicitly quantify previous results showing higher conservation of active promoter regions compared to distal enhancers in selected representatives of mammals (Xiao et al., 2012) or within primates (Cotney et al., 2013).

Our study has a number of limitations. First, the relationship between different histone marks and the activity of enhancers is not perfectly understood. Most active enhancers are marked

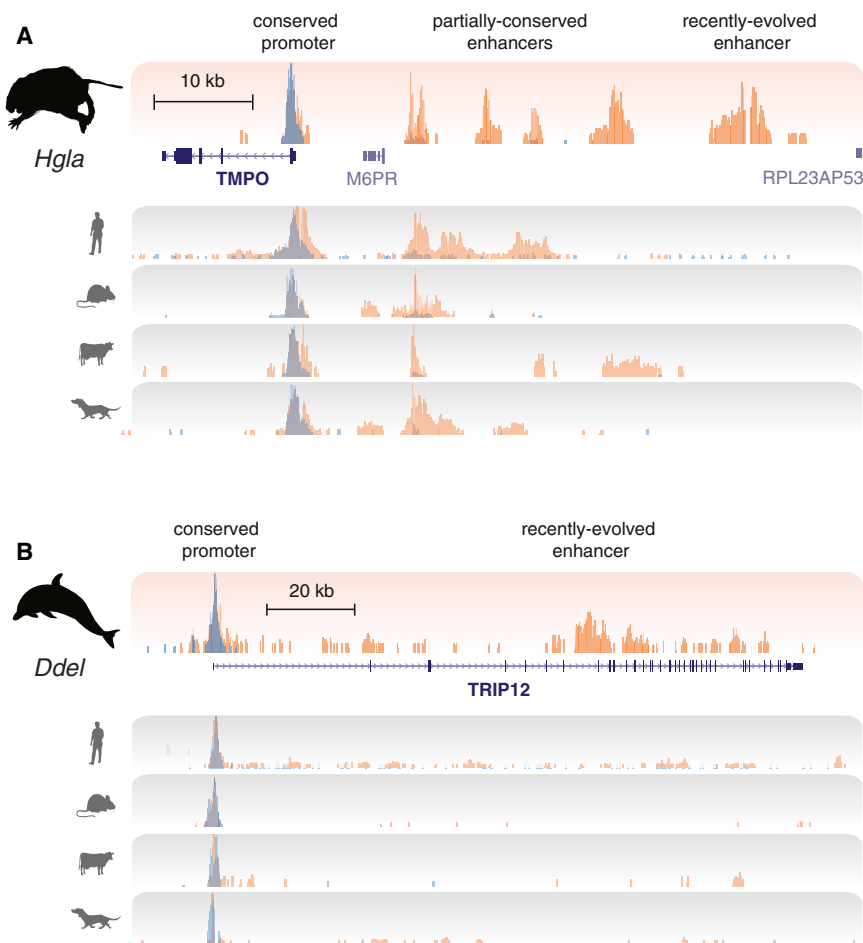


Figure 7. Recently Evolved Enhancers Associate with Genes under Positive Selection during Naked Mole Rat and Dolphin Evolution

(A) The liver enhancer and promoter landscape surrounding the *TMPO* locus, which is under positive selection in naked mole rat (Kim et al., 2011), is shown (upper track). The bottom four tracks display overlaid H3K4me3 (blue) and H3K27ac (orange) levels in the orthologous regions of human, mouse, dog, and cow. Shown (left to right) are a promoter present in all species, four enhancer regions shared in a subset of species, and a naked mole rat-specific enhancer whose recently evolved activity is not present in other study species.

(B) The enhancer and promoter landscape surrounding the *TRIP12* locus, which is under positive selection in dolphins (Sun et al., 2013), is shown. In this case, no mammals other than dolphin show liver enhancer activity near this gene; this enhancer is thus a good candidate to contain the regulatory regions associated with positive selection in dolphin.

See also Table S5.

by H3K27ac (Andersson et al., 2014; Creighton et al., 2010; Zhu et al., 2013), and typically over two-thirds of regions enriched for H3K27ac show independent evidence in transgenesis assays for regulatory activity (Nord et al., 2013). Global mapping of H3K4me1 and p300 can also detect poised enhancer activity genome-wide, which can partly differ from that identified by H3K27ac (Heintzman et al., 2007; Krebs et al., 2011; Visel et al., 2009). Second, other approaches to map regulatory sequences, such as DNase-seq (Shibata et al., 2012) or ATAC-seq (Buenrostro et al., 2013), can reveal all regions of open chromatin genome-wide, but cannot distinguish promoters and enhancers. Third, our approach does not directly reveal which transcription factors control these regulatory regions, as would a more direct comparison (Kunarso et al., 2010; Paris et al., 2013; Schmidt et al., 2010), which in turn can only capture a modest subset of active regions. Fourth, our results generalize to other mammalian somatic tissues to the extent that adult liver is a representative tissue. However, other studies have suggested rapid enhancer evolution in mammals, using embryonic limb buds (Cotney et al., 2013), adipocytes (Mikkelsen et al., 2010), and embryonic stem cells (Xiao et al., 2012). These studies and others (Barbosa-Morais et al., 2012; Brawand et al., 2011) suggest that regulation in other somatic tissues evolves similarly,

though embryonic tissues and their enhancers may be under stronger evolutionary constraint (Faure et al., 2012; He et al., 2011; Nord et al., 2013). Fifth, we cannot directly evaluate how often regions with regulatory activity are fully tissue-specific, particularly among those we assign as enhancers (Zhu et al., 2013).

One powerful strategy to dissect the regulatory genome has been to identify

regions under high sequence constraint (Lindblad-Toh et al., 2011). Testing for activity has revealed that thousands of constrained noncoding regulatory sequences can act as enhancers in embryonic tissues (Pennacchio et al., 2006). The complementary approach we used additionally captures rapidly evolving regulatory regions. The enhancer regions we mapped likely range in function from essential to dispensable, which is reflected both in the modest sequence constraint and rapid evolution between species. Most of these regions would likely be missed by any sequence-conservation based approach. On the other hand, many DNA sequences we do not identify as enhancers may be active in other tissues or embryonic states, which we anticipate to be an area of active investigation.

Rapid enhancer and slow promoter evolution is a fundamental property of the mammalian regulatory genome. Active enhancer elements have a mean lifetime three times shorter than active promoters do, despite similar alignability of their underlying DNA sequences. Comparative sequence-based approaches have limited power to detect regulatory regions, in part because of their rapid evolution (Alföldi and Lindblad-Toh, 2013; Lindblad-Toh et al., 2011); indeed, our data indicate that sequence-based features such as sequence constraint or TF binding site density are poor predictors of enhancer

conservation. Nevertheless, previous work across *Drosophila* species has indicated that specific TF motifs may be preferentially preserved in functionally conserved enhancers (Arnold et al., 2014). In agreement, we found motifs for the liver-specific transcription factor CEBPA enriched in highly conserved liver enhancers.

Active Mammalian Enhancers Are Predominantly Apomorphic

Our results also newly reveal thousands of functionally active regulatory regions conserved across placental mammals, the vast majority of which are proximal promoter sequences. Placental-conserved proximal promoters in mammalian liver are commonly associated with ubiquitously expressed genes. In contrast, only 12% of highly conserved regulatory regions are active enhancers and these are near genes associated with liver-specific activities.

Perhaps our most surprising finding is that representative mammals typically deploy over 10,000 enhancers in a lineage- and probably most often species-specific manner. In total, almost half of all enhancers in each species appear to be recently evolved. Our results confirm and extend the concept that exaptation is a widespread phenomenon across placental mammals (Cotney et al., 2013), and redeployment of ancestral DNA is the dominant mechanism to generate active enhancers across a diverse cross-section of mammals. Interestingly, a recent study comparing enhancer activity across the much smaller genomes of five *Drosophila* species (Arnold et al., 2014) found a similar proportion of gained enhancers, especially for more distant species.

Another mechanism to create regulatory sequences is repeat-carried expansion of regulatory elements. Recent studies have indicated the involvement of specific repeat element expansions in the de novo creation of TF binding sites for CTCF (Bourque et al., 2008; Schmidt et al., 2012), Oct4/Nanog (Kunarto et al., 2010), and NRSF (Mortazavi et al., 2006). Our results show that repeat-carriage of newborn enhancers is not the dominant evolutionary process in mammals: repeat element enrichment is only significant among the recently evolved enhancers found in DNA less than 40 Ma old. Two technical limitations may have caused us to underestimate the repeat-driven creation of recently evolved enhancers (also, see Jacques et al., 2013): the difficulty of mapping reads to recently duplicated regions, and the incomplete representation of repeat regions in genome assemblies.

Recently Evolved Promoters, Though Less Common Than Enhancers, Are Mostly Found in Young DNA

Promoters are far more evolutionarily stable than are enhancers. Nevertheless, the absolute number of promoters deeply conserved across all 20 study species is similar to the number of recently evolved promoters in any one species. Compared to the tens of thousands of newborn enhancers arising from exaptation of ancestral DNA, there are few newborn promoters—and these often arise from DNA sequences that are themselves evolutionarily young. We were not able to identify sequence features that account for the birth of promoters in young DNA. In contrast, the recently evolved promoters arising in ancestral sequences overlap LTR repeats, which enrich for latent non-coding RNA activity (Fort et al., 2014).

A Strategy for Identifying the Enhancer Repertoire of Unannotated Genomes

Finally, extending an approach pioneered in well-annotated primate genomes (Cotney et al., 2013; Shibata et al., 2012), we provide examples of how experimental mapping of enhancers and promoters in newly sequenced mammals can annotate the regulatory network of genes, which have been identified computationally as under positive selection. Across representative species, we discovered that recently evolved enhancers are significantly over-represented in the vicinity of positively-selected genes and can often suggest candidate regulatory elements that could mediate species-specific adaptations. This result was obtained using only a single somatic tissue. Similarly, significant associations likely also exist in between the newly evolved enhancers specific to other somatic tissues and positively selected genes, which would uncover an extensive repertoire of highly evolvable, potentially synergistic regulatory connections.

Future Directions

Our quantitation and analysis of the evolution of promoters and enhancers across a wide cross-section of mammals has revealed how dynamic and rapid enhancer evolution is. Within this regulatory diversity are the instructions by which a small number of founder species have radiated into surprising new niches, including marine (cetaceans) and aerial environments (bats). By combining detailed investigations of carefully selected subclades with new tools for modifying any sequenced genome, future studies will identify, formalize, and explore the functional instructions directing the diversity of mammalian forms.

EXPERIMENTAL PROCEDURES

We performed ChIP-seq using liver tissue isolated from 20 mammalian species (Table S1). At least two independent biological replicates from different animals, generally young adult males, were performed for each species and antibody. The only exception was *Balaenoptera borealis*, for which a single individual was profiled, and dolphin, for which we profiled a single individual from two closely-related species. ChIP-seq experiments were performed as recently described (Aldridge et al., 2013) with antibodies against H3K4me3 (Millipore 05-1339) and H3K27ac (Abcam ab4729). To match inter-individual variability for the two histone marks, the same tissue samples were used for both antibodies and control input DNA in each species.

Sequencing reads were aligned to the appropriate reference genome with BWA v.0.5.9 (Table S2) and regions of enrichment determined with MACS v1.4.2. Regions enriched in two to four biological replicates and overlapping by a minimum 50% of their length were merged and categorized into active promoters (H3K4me3-enriched regions, with or without overlapping H3K27ac enrichment) or enhancers (regions enriched only for H3K27ac). Cross-species comparisons were performed through the Ensembl API. Human, macaque, vervet, marmoset, mouse, rat, rabbit, cow, pig, dog, and cat were directly cross-compared using the 13 eutherian mammals EPO alignment available from Ensembl (Flicek et al., 2014). Species not included in the EPO alignment were compared to the reference species of their respective clade (human, mouse, cow, dog, or opossum) using Lastz alignments. Promoters or enhancers were considered as having conserved activity between species when their orthologous location in the second species overlapped a marked region by a minimum of 50% in length. All pairwise comparisons correspond to average values of reciprocal comparisons between species. Genome annotations (including gene ontology and repetitive and constrained elements) were downloaded from Ensembl v73. See also [Extended Experimental Procedures](#).

ACCESSION NUMBERS

Data have been deposited under ArrayExpress accession number E-MTAB-2633.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and seven tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.01.006>.

AUTHOR CONTRIBUTIONS

D.V., C.B., P.F., and D.T.O. designed experiments; D.V. and S.A. performed experiments; C.B., D.V., T.F.R., and M.L. analyzed the data; T.J.P., R.D., J.T.E., A.J.J., J.M.A.T., M.F.B., and E.P.M. provided tissue samples; M.P. generated LastZ whole-genome alignments; D.V., C.B., P.F., and D.T.O. wrote the manuscript; P.F. and D.T.O. oversaw the work. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We thank Stephen Watt, Frances Connor, the CRUK-CI Genomics and Bioinformatics cores, Biological Resources Unit (Matthew Clayton), Margaret Brown (West Yorkshire bat hospital), Julie E. Horvath (North Carolina Central University), and Chris Dillingham (University of Cardiff) for technical assistance; Matthieu Muffato for assistance with whole-genome alignments; Claudia Kutter, Gordon Brown, Christine Feig, and Christina Ernst for useful comments and discussions, and the EBI systems team for management of computational resources. This research was supported by Cancer Research UK (D.V., D.T.O.), the European Molecular Biology Laboratory (C.B., P.F.), the Wellcome Trust (WT095908) (P.F.) and (WT098051) (P.F., D.T.O.), the European Research Council, EMBO Young Investigator Programme (D.T.O.), the National Science Foundation (0744979) (T.J.P.), NIH (P40 OD010965, R01 OD010980, R37 MH060233) (A.J.J.) and MRC (U117588498) (J.M.A.T.). Cetacean samples were collected by the UK Cetacean Strandings Investigation Programme, funded by Defra and the Governments of Scotland and Wales.

Received: August 29, 2014

Revised: October 31, 2014

Accepted: December 15, 2014

Published: January 29, 2015

REFERENCES

- Aldridge, S., Watt, S., Quail, M.A., Rayner, T., Lusk, M., Bimson, M.F., Gaffney, D., and Odom, D.T. (2013). AHT-ChIP-seq: a completely automated robotic protocol for high-throughput chromatin immunoprecipitation. *Genome Biol.* 14, R124.
- Alföldi, J., and Lindblad-Toh, K. (2013). Comparative genomics as a tool to understand evolution and disease. *Genome Res.* 23, 1063–1068.
- Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al.; FANTOM Consortium (2014). An atlas of active enhancers across human cell types and tissues. *Nature* 507, 455–461.
- Arnold, C.D., Gerlach, D., Spies, D., Matts, J.A., Sytnikova, Y.A., Pagani, M., Lau, N.C., and Stark, A. (2014). Quantitative genome-wide enhancer activity maps for five *Drosophila* species show functional enhancer conservation and turnover during cis-regulatory evolution. *Nat. Genet.* 46, 685–692.
- Ballester, B., Medina-Rivera, A., Schmidt, D., González-Porta, M., Carlucci, M., Chen, X., Chessman, K., Faure, A.J., Funnell, A.P., Gonçalves, A., et al. (2014). Multi-species, multi-transcription factor binding highlights conserved control of tissue-specific biological pathways. *eLife* 3, e02626.
- Barbosa-Morais, N.L., Irimia, M., Pan, Q., Xiong, H.Y., Gueroussov, S., Lee, L.J., Slobodeniuc, V., Kutter, C., Watt, S., Colak, R., et al. (2012). The evolutionary landscape of alternative splicing in vertebrate species. *Science* 338, 1587–1593.
- Bourque, G., Leong, B., Vega, V.B., Chen, X., Lee, Y.L., Srinivasan, K.G., Chew, J.L., Ruan, Y., Wei, C.L., Ng, H.H., and Liu, E.T. (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res.* 18, 1752–1762.
- Brawand, D., Soumillon, M., Necsulea, A., Julien, P., Csárdi, G., Harrigan, P., Weier, M., Liechti, A., Aximu-Petri, A., Kircher, M., et al. (2011). The evolution of gene expression levels in mammalian organs. *Nature* 478, 343–348.
- Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* 10, 1213–1218.
- Cain, C.E., Blekhman, R., Marioni, J.C., and Gilad, Y. (2011). Gene expression differences among primates are associated with changes in a histone epigenetic modification. *Genetics* 187, 1225–1234.
- Chan, E.T., Quon, G.T., Chua, G., Babak, T., Trochesset, M., Zirngibl, R.A., Aubin, J., Ratcliffe, M.J., Wilde, A., Brudno, M., et al. (2009). Conservation of core gene expression in vertebrate tissues. *J. Biol.* 8, 33.
- Cotney, J., Leng, J., Yin, J., Reilly, S.K., DeMare, L.E., Emera, D., Ayoub, A.E., Rakic, P., and Noonan, J.P. (2013). The evolution of lineage-specific regulatory activities in the human embryonic limb. *Cell* 154, 185–196.
- Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. USA* 107, 21931–21936.
- Degner, J.F., Pai, A.A., Pique-Regi, R., Veyrieras, J.B., Gaffney, D.J., Pickrell, J.K., De Leon, S., Michelini, K., Lewellen, N., Crawford, G.E., et al. (2012). DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature* 482, 390–394.
- Faure, A.J., Schmidt, D., Watt, S., Schwalie, P.C., Wilson, M.D., Xu, H., Ramsay, R.G., Odom, D.T., and Flícek, P. (2012). Cohesin regulates tissue-specific expression by stabilizing highly occupied cis-regulatory modules. *Genome Res.* 22, 2163–2175.
- Flícek, P., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., et al. (2014). Ensembl 2014. *Nucleic Acids Res.* 42, D749–D755.
- Forrest, A.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J., Haberle, V., Lassman, T., Kulakovskiy, I.V., Lizio, M., Itoh, M., et al.; FANTOM Consortium and the RIKEN PMI and CLST (DGT) (2014). A promoter-level mammalian expression atlas. *Nature* 507, 462–470.
- Fort, A., Hashimoto, K., Yamada, D., Salimullah, M., Keya, C.A., Saxena, A., Bonetti, A., Voineagu, I., Bertin, N., Kratz, A., et al.; FANTOM Consortium (2014). Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nat. Genet.* 46, 558–566.
- Frith, M.C., Ponjavic, J., Fredman, D., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y., and Sandelin, A. (2006). Evolutionary turnover of mammalian transcription start sites. *Genome Res.* 16, 713–722.
- Hare, E.E., Peterson, B.K., Iyer, V.N., Meier, R., and Eisen, M.B. (2008). Sepsid even-skipped enhancers are functionally conserved in *Drosophila* despite lack of sequence conservation. *PLoS Genet.* 4, e1000106.
- He, Q., Bardet, A.F., Patton, B., Purvis, J., Johnston, J., Paulson, A., Gogol, M., Stark, A., and Zeitlinger, J. (2011). High conservation of transcription factor binding and evidence for combinatorial regulation across six *Drosophila* species. *Nat. Genet.* 43, 414–420.
- Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* 39, 311–318.
- Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone

- modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459, 108–112.
- Jacques, P.E., Jeyakani, J., and Bourque, G. (2013). The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genet.* 9, e1003504.
- Kheradpour, P., Ernst, J., Melnikov, A., Rogov, P., Wang, L., Zhang, X., Alston, J., Mikkelsen, T.S., and Kellis, M. (2013). Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res.* 23, 800–811.
- Kim, J., He, X., and Sinha, S. (2009). Evolution of regulatory sequences in 12 *Drosophila* species. *PLoS Genet.* 5, e1000330.
- Kim, E.B., Fang, X., Fushan, A.A., Huang, Z., Lobanov, A.V., Han, L., Marino, S.M., Sun, X., Turanov, A.A., Yang, P., et al. (2011). Genome sequencing reveals insights into physiology and longevity of the naked mole rat. *Nature* 479, 223–227.
- Krebs, A.R., Karmodiya, K., Lindahl-Alten, M., Struhl, K., and Tora, L. (2011). SAGA and ATAC histone acetyl transferase complexes regulate distinct sets of genes and ATAC defines a class of p300-independent enhancers. *Mol. Cell* 44, 410–423.
- Kunarso, G., Chia, N.Y., Jeyakani, J., Hwang, C., Lu, X., Chan, Y.S., Ng, H.H., and Bourque, G. (2010). Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat. Genet.* 42, 631–634.
- Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M.F., Parker, B.J., Washietl, S., Kheradpour, P., Ernst, J., Jordan, G., Mauceli, E., et al.; Broad Institute Sequencing Platform and Whole Genome Assembly Team; Baylor College of Medicine Human Genome Sequencing Center Sequencing Team; Genome Institute at Washington University (2011). A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 478, 476–482.
- Ludwig, M.Z., Palsson, A., Alekseeva, E., Bergman, C.M., Nathan, J., and Kreitman, M. (2005). Functional evolution of a cis-regulatory module. *PLoS Biol.* 3, e93.
- Lynch, M., Bobay, L.M., Catania, F., Gout, J.F., and Rho, M. (2011). The repatterning of eukaryotic genomes by random genetic drift. *Annu. Rev. Genomics Hum. Genet.* 12, 347–366.
- McLean, C.Y., Reno, P.L., Pollen, A.A., Bassan, A.I., Capellini, T.D., Guenther, C., Indjeian, V.B., Lim, X., Menke, D.B., Schaar, B.T., et al. (2011). Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature* 471, 216–219.
- Merkin, J., Russell, C., Chen, P., and Burge, C.B. (2012). Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science* 338, 1593–1599.
- Mikkelsen, T.S., Xu, Z., Zhang, X., Wang, L., Gimble, J.M., Lander, E.S., and Rosen, E.D. (2010). Comparative epigenomic analysis of murine and human adipogenesis. *Cell* 143, 156–169.
- Mortazavi, A., Leeper Thompson, E.C., Garcia, S.T., Myers, R.M., and Wold, B. (2006). Comparative genomics modeling of the NRSF/REST repressor network: from single conserved sites to genome-wide repertoire. *Genome Res.* 16, 1208–1221.
- Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C., and Clark, A.G. (2007). Recent and ongoing selection in the human genome. *Nat. Rev. Genet.* 8, 857–868.
- Nord, A.S., Blow, M.J., Attanasio, C., Akiyama, J.A., Holt, A., Hosseini, R., Phouanavong, S., Plajzer-Frick, I., Shoukry, M., Afzal, V., et al. (2013). Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell* 155, 1521–1531.
- Paris, M., Kaplan, T., Li, X.Y., Villalta, J.E., Lott, S.E., and Eisen, M.B. (2013). Extensive divergence of transcription factor binding in *Drosophila* embryos with highly conserved gene expression. *PLoS Genet.* 9, e1003748.
- Pennacchio, L.A., Ahituv, N., Moses, A.M., Prabhakar, S., Nobrega, M.A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K.D., et al. (2006). In vivo enhancer analysis of human conserved non-coding sequences. *Nature* 444, 499–502.
- Pollard, K.S., Salama, S.R., Lambert, N., Lambot, M.A., Coppens, S., Pedersen, J.S., Katzman, S., King, B., Onodera, C., Siepel, A., et al. (2006). An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* 443, 167–172.
- Prabhakar, S., Poulin, F., Shoukry, M., Afzal, V., Rubin, E.M., Couronne, O., and Pennacchio, L.A. (2006). Close sequence comparisons are sufficient to identify human cis-regulatory elements. *Genome Res.* 16, 855–863.
- Rands, C.M., Meader, S., Ponting, C.P., and Lunter, G. (2014). 8.2% of the Human genome is constrained: variation in rates of turnover across functional element classes in the human lineage. *PLoS Genet.* 10, e1004525.
- Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C., Schreiber, S.L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. *Nature* 419, 407–411.
- Schmidt, D., Wilson, M.D., Ballester, B., Schwalie, P.C., Brown, G.D., Marshall, A., Kutter, C., Watt, S., Martinez-Jimenez, C.P., Mackay, S., et al. (2010). Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* 328, 1036–1040.
- Schmidt, D., Schwalie, P.C., Wilson, M.D., Ballester, B., Gonçalves, A., Kutter, C., Brown, G.D., Marshall, A., Flicek, P., and Odom, D.T. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* 148, 335–348.
- Schwaiger, M., Schöner, A., Rendeiro, A.F., Pribitzer, C., Schauer, A., Gilles, A.F., Schinko, J.B., Renfer, E., Fredman, D., and Technau, U. (2014). Evolutionary conservation of the eumetazoan gene regulatory landscape. *Genome Res.* 24, 639–650.
- Shibata, Y., Sheffield, N.C., Fedrigo, O., Babbitt, C.C., Wortham, M., Tewari, A.K., London, D., Song, L., Lee, B.K., Iyer, V.R., et al. (2012). Extensive evolutionary changes in regulatory element activity during human origins are associated with altered gene expression and positive selection. *PLoS Genet.* 8, e1002789.
- Stefflova, K., Thybert, D., Wilson, M.D., Streeter, I., Aleksic, J., Karagianni, P., Brazma, A., Adams, D.J., Talianidis, I., Marioni, J.C., et al. (2013). Cooperativity and rapid evolution of cobound transcription factors in closely related mammals. *Cell* 154, 530–540.
- Sun, Y.B., Zhou, W.P., Liu, H.Q., Irwin, D.M., Shen, Y.Y., and Zhang, Y.P. (2013). Genome-wide scans for candidate genes involved in the aquatic adaptation of dolphins. *Genome Biol. Evol.* 5, 130–139.
- Villar, D., Flicek, P., and Odom, D.T. (2014). Evolution of transcription factor binding in metazoans - mechanisms and functional implications. *Nat. Rev. Genet.* 15, 221–233.
- Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457, 854–858.
- Wray, G.A. (2007). The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* 8, 206–216.
- Xiao, S., Xie, D., Cao, X., Yu, P., Xing, X., Chen, C.C., Musselman, M., Xie, M., West, F.D., Lewin, H.A., et al. (2012). Comparative epigenomic annotation of regulatory DNA. *Cell* 149, 1381–1392.
- Zhu, J., Adli, M., Zou, J.Y., Verstappen, G., Coyne, M., Zhang, X., Durham, T., Miri, M., Deshpande, V., De Jager, P.L., et al. (2013). Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell* 152, 642–654.

Acetate Fuels the Cancer Engine

Costas A. Lyssiotis* and Lewis C. Cantley*

*Correspondence: col2007@med.cornell.edu (C.A.L.), lcantley@med.cornell.edu (L.C.C.)
<http://dx.doi.org/10.1016/j.cell.2015.01.021>

(Cell 159, 1492–1494; December 18, 2014)

Due to a production error, a label in Figure 1 of the Preview above was incorrect. Within the figure, the text “ADP, P_i” should have read “AMP, PP_i.” The figure has been corrected online and appears below.

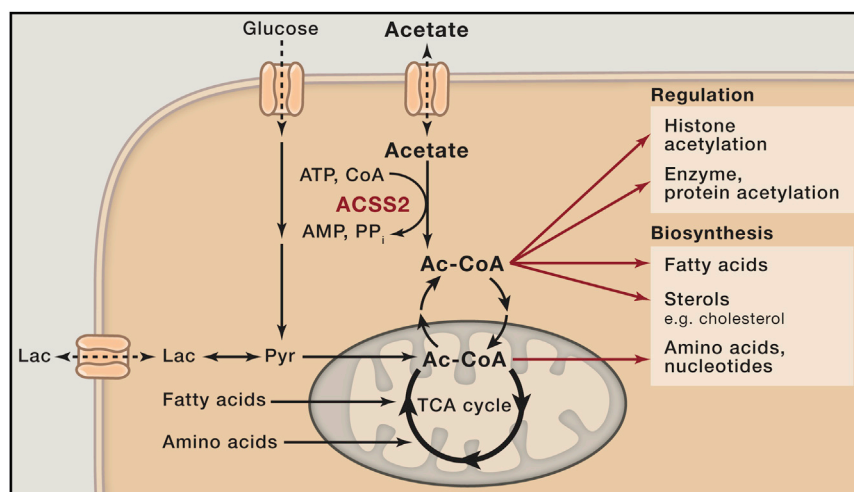


Figure 1. Acetyl-CoA Is a Central Node in Carbon Metabolism

ATRX Directs Binding of PRC2 to Xist RNA and Polycomb Targets

Kavitha Sarma, Catherine Cifuentes-Rojas, Ayla Ergun, Amanda del Rosario, Yesu Jeon, Forest White, Ruslan Sadreyev, and Jeannie T. Lee*

*Correspondence: lee@molbio.mgh.harvard.edu
<http://dx.doi.org/10.1016/j.cell.2015.01.030>

(Cell 159, 869–883; November 6, 2014)

In characterizing ATRX as a regulator of X inactivation, we used two shRNAs to knockdown ATRX and then visualized the effects on Xist localization using FISH. In Figure 1C, we erroneously presented duplicate images of cells treated with shATRX2 showing both Xist and Xist plus DAPI staining in place of images showing cells treated with shATRX1. The quantification of Xist levels presented in Figure 1C reflects a separate qRT-PCR analysis and is not affected. The corrected figure, showing shATRX1-treated cells from the same experiment, appears below. Additionally, in the interest of experimental transparency, the accompanying figure legend has been updated to include the data acquisition parameters used in the experiment from which the images were derived. The figure and figure legend have been corrected online.

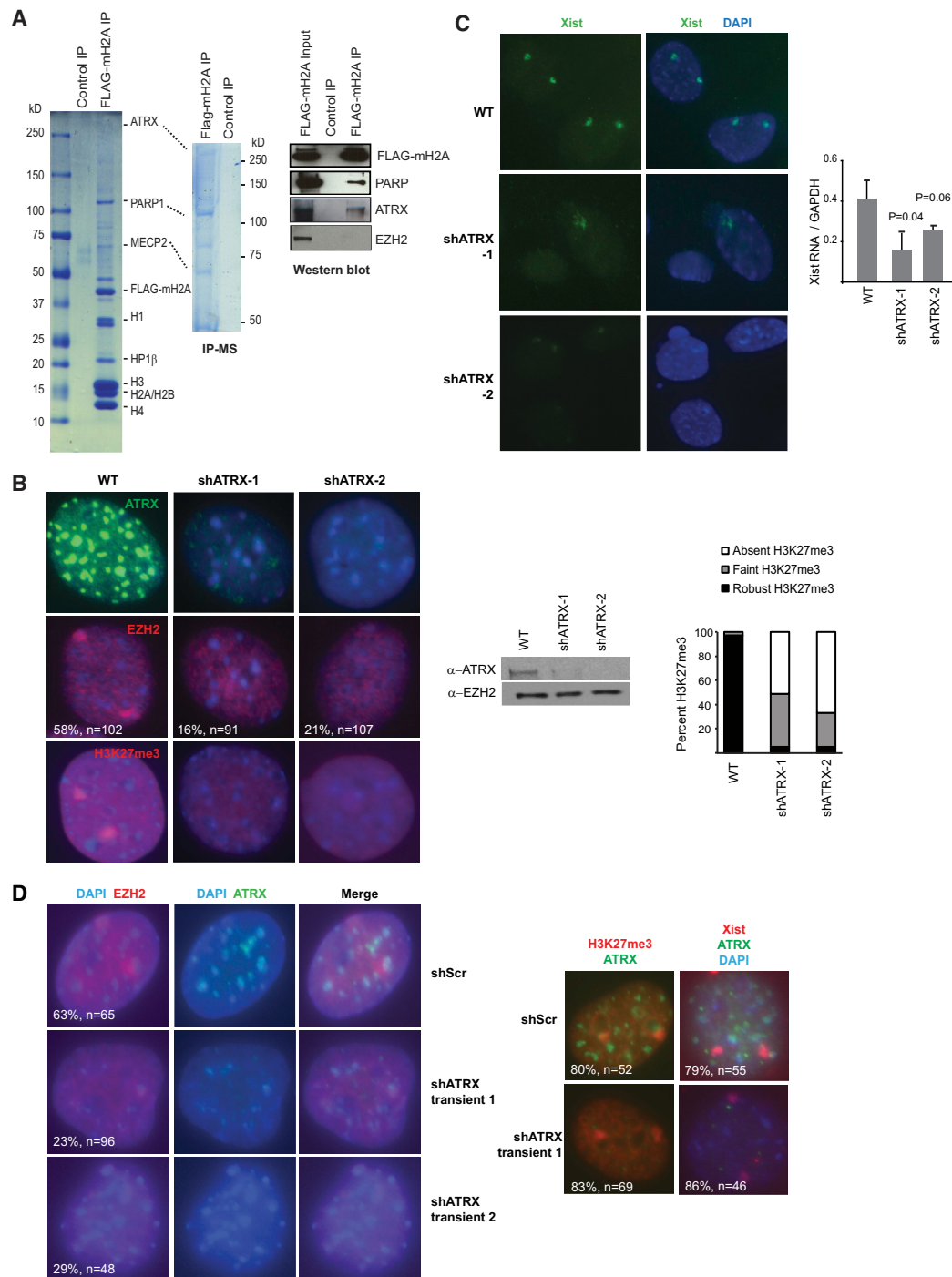


Figure 1. A Proteomics Screen Identifies ATRX as a Candidate XCI Regulator

(A) IP-MS: Colloidal blue staining of FLAG IP from control (293F) and FLAG-mH2A-expressing 293 run on a 4%–20% (left) and a 6% (right) SDS gradient gel. FLAG IP was validated by western blot.

(B) Left: Immunostaining of ATRX, EZH2, and H3K27me3 in WT and two independent stable ATRX-KD MEF lines (shATRX-1,-2). Sample size (n) and %EZH2 association with Xi are shown. Middle: western blot showing ATRX depletion but constant EZH2 levels in shATRX-1 and shATRX-2 female MEFs. Right: Patterns of H3K27me3 observed. n = 100–150 per experiment.

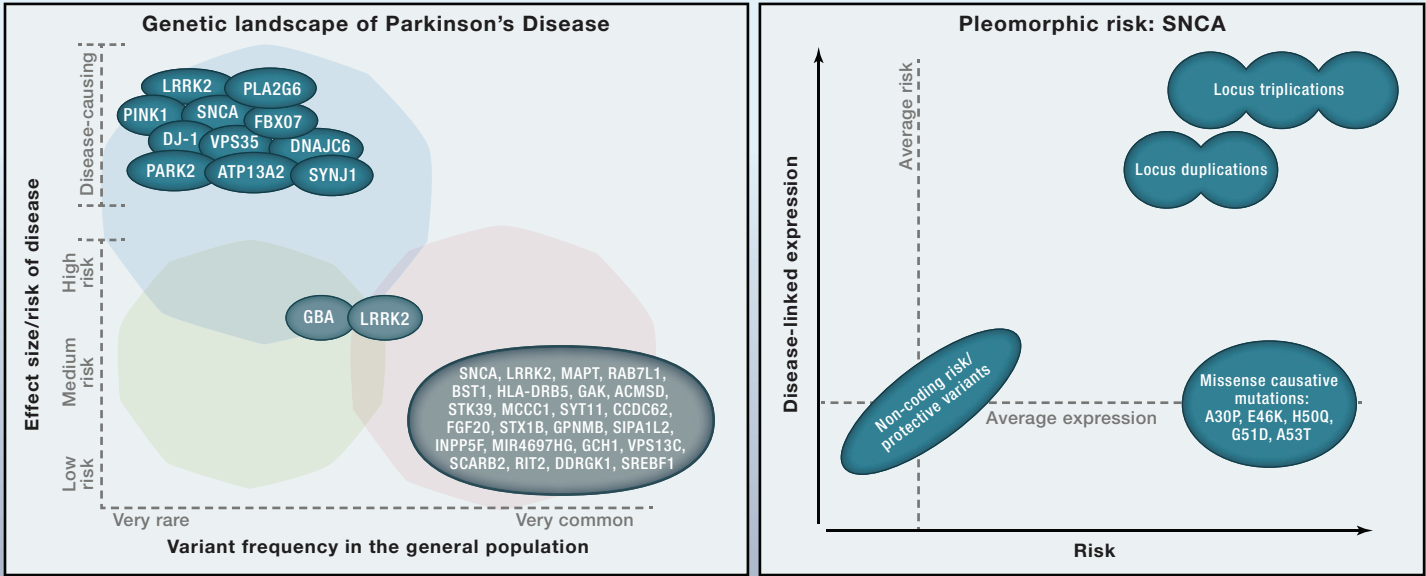
(C) Left: Xist RNA FISH in indicated fibroblast lines with FITC acquisition times of 500 ms and a gain of 61 for all samples. Right: qRT-PCR analyses of Xist RNA levels. SE bars from three independent experiments are shown with Student's t test p values.

(D) Left: Immunostaining of ATRX and EZH2 in MEFs transiently transfected with scrambled shRNA (shScr) and two shATRX constructs (shATRX-1, shATRX-2). Right: H3K27me3 staining and Xist RNA FISH show no change in the intensity or foci number after transient ATRX KD.

SnapShot: Genetics of Parkinson's Disease

Cell

José Brás, Rita Guerreiro, and John Hardy
Department of Molecular Neuroscience, Institute of Neurology, University College London, Queen Square,
London WC1N 1PJ, UK



Gene official symbol	Gene name	Location	Possible pathways / pathological biological processes
MENDELIAN GENES			
SNCA	Synuclein, alpha	4q21	Synaptic function; mitochondrial function; autophagy/lysosomal degradation
PARK2	Parkin RBR E3 ubiquitin protein ligase	6q25.2-q27	Mitochondrial function/mitophagy; ubiquitination; synaptic function
PINK1	PTEN -induced putative kinase 1	1p36	Mitochondrial function/mitophagy
PARK7/DJ-1	Parkinson protein 7	1p36.23	Inflammation/immune system; mitochondrial function
LRRK2	Leucine-rich repeat kinase 2	12q12	Synaptic function; inflammation/immune system; autophagy/lysosomal degradation
PLA2G6	Phospholipase A2, group VI	22q13.1	Mitochondrial function
FBX07	F-box protein 7	22q12.3	Ubiquitination; mitochondrial function/mitophagy
VPS35	Vacuolar protein sorting 35 homolog (S. cerevisiae)	16q12	Autophagy/lysosomal degradation; endocytosis
ATP13A2	ATPase type 13A2	1p36	Mitochondrial function; autophagy/lysosomal degradation
DNAJC6	DnaJ (Hsp40) homolog, subfamily C, member 6	1p31.3	Synaptic function; endocytosis
SYNJ1	Synaptojanin 1	21q22.2	Synaptic function; endocytosis
RISK GENES			
GBA	Glucosidase, beta, acid	1q21	Inflammation/immune system; autophagy/lysosomal degradation; metabolic pathways
RISK LOCI			
MAPT	Microtubule-associated protein tau	17q21.1	Microtubule stabilization and axonal transport
RAB7L1	RAB7, member RAS oncogene family-like 1	1q32	Autophagy/lysosomal degradation
BST1	Bone marrow stromal cell antigen 1	4p15	Immune system
HLA-DRB5	Major histocompatibility complex, class II, DR beta 5	6p21.3	Inflammation/immune system
GAK	Cyclin-G-associated kinase	4p16	Autophagy/lysosomal degradation; synaptic function; endocytosis
ACMSD	Aminocarboxymuconate semialdehyde decarboxylase	2q21.3	Tryptophan metabolism; metal ion binding; metabolic pathways
STK39	Serine threonine kinase 39	2q24.3	Inflammation/immune system; protein kinase binding; cellular stress response
SYT11	Synaptotagmin XI	1q21.2	Synaptic function; transporter activity; metal ion binding; substrate for PARK2
FGF20	Fibroblast growth factor 20	8p22	Growth factor activity; FGF receptor binding
STX1B	Syntaxin 1B	16p11.2	Synaptic function; SNAP receptor activity; protein domain-specific binding
GPNMB	Glycoprotein (transmembrane) nmb	7p15	Integrin binding; heparin binding; cancer pathways
SIPA1L2	Signal-induced proliferation-associated 1 like 2	1q42.2	GTPase activator activity
INPP5F	Inositol polyphosphate-5-phosphatase F	10q26.11	Phosphoric ester hydrolase activity
MIR4697HG	MIR4697 host gene (non-protein coding)	11q25	
GCH1	GTP cyclohydrolase 1	14q22.1-q22.2	GTP binding; calcium ion binding; BH4 metab; metabolic pathways
VPS13C	Vacuolar protein sorting 13 homolog C (S. cerevisiae)	15q22.2	Endocytosis
DDRGK1	DDRGK domain containing 1	20p13	Protein binding
MCCC1	Methylcrotonoyl-CoA carboxylase 1 (alpha)	3q27	Biotin carboxylase activity; methylcrotonoyl-CoA carboxylase activity; metabolic pathways
SCARB2	Scavenger receptor class B, member 2	4q21.1	Autophagy/lysosomal degradation; receptor activity (lysosomal receptor for GBA targeting); enzyme binding
CCDC62	Coiled-coil domain containing 62	12q24.31	Nuclear receptor coactivator; cancer pathways
RIT2	Ras-like without CAAX 2	18q12.3	Synaptic function; calmodulin binding; GTP binding
SREBF1	Sterol regulatory element binding transcription factor 1	17p11.2	Chromatin binding; cholesterol and steroid metabolic processes

SnapShot: Genetics of Parkinson's Disease

Cell

José Brás, Rita Guerreiro, and John Hardy

Department of Molecular Neuroscience, Institute of Neurology, University College London, Queen Square, London WC1N 1PJ, UK

Different types of genetic technologies and approaches allow for the study and identification of different types of genetic variability in a disease. Here, represented are the genes and genetic loci independently replicated as being associated with the development of Parkinson's disease (PD)/parkinsonism.

Genetic analyses of familial cases (mainly genetic linkage [blue area] and, more recently, exome sequencing [green area]) have led to the identification of causative mutations in 11 genes implicated in monogenic typical or atypical forms of parkinsonism. From these, eight genes have been associated with autosomal-recessive patterns of inheritance, either causing typical early-onset PD (*PARK2*, *PINK1*, and *DJ-1/PARK7*) or atypical forms of parkinsonism with juvenile onsets (*ATP13A2*, *PLA2G6*, *FBXO7*, *DNAJC6*, and *SYNJ1*). Three genes have been shown to cause typical autosomal dominant PD phenotypes (*SNCA*, *LRRK2*, and *VPS35*) associated with early- or late-onsets of disease. Additionally, other genes are known to harbor mutations associated with non-PD disorders that may present with parkinsonism, for example *ATXN2*, *ATXN3*, *GCH1*, *GRN*, *MAPT*, *C9ORF72*, *CSF1R*, *TH*, and *SPG1*. Very recently, mutations in *RAB39B* have been described as causing X-linked intellectual disability plus a phenotype indistinguishable from early-onset PD.

Other chromosomal loci (like *PARK3* and *PARK10*, for example) have been identified by genome-wide approaches as genomic regions associated with typical PD. These loci may contain other, still-to-be-identified, PD genes.

Lastly, there are also some genes that have been suggested to harbor causative PD mutations, which have not been confirmed: *GIGYF2*, *HTRA2*, *UCHL1*, *EIF4G1*, and *SPR*.

By using genetic family studies (for *LRRK2*) and a candidate gene approach (for *GBA*), high-risk variants (with odds ratio in the range of 5–8) for the development of typical PD have been identified in both genes (central part of the graph).

More recently, the development of whole-genome genotyping platforms (pink area of the graph) has allowed for the study of the involvement of common variants with low risk in the disease. This has led to the identification of 24 new genetic loci by several independent genome-wide association studies (GWAS) and meta-analyses: *SNCA*, *LRRK2*, *MAPT*, *RAB7L1*, *BST1*, *HLA-DRB5*, *GAK*, *ACMSD*, *STK39*, *MCCC1*, *SYT11*, *CCDC62*, *FGF20*, *STX1B*, *GPMB*, *SIPA1L2*, *INPP5F*, *MIR4697HG*, *GCH1*, *VPS13C*, *SCARB2*, *RIT2*, *DDRGK1*, and *SREBF1*. Because of the way these studies are designed, they only identify genetic regions associated with disease and not specific genes or variants. For this reason, if the significant single nucleotide polymorphisms (SNPs) are intergenic or the region contains more than one gene, the locus usually gets its name from the gene closest to the significant hit. Very few of these significant hits have a clear functional role in the disease and, because of this, follow-up work is currently underway to determine exactly which genes and genetic variants are important for the disease and how they are exerting their effect. Recently, an unbiased screen for interactors of *LRRK2* identified the most likely candidates for two of these GWAS loci: in chromosome 1q32, the associated locus (originally named *RAB7L1/NUCKS1*) contained five genes, and in chromosome 4p16, the associated locus (originally named *GAK/TMEM175/DGKQ*) contained nearly ten genes. *RAB7L1* and *GAK* have now been identified as *LRRK2* interactors and, in this way, as the most likely hits in each region. Additionally, these proteins were shown to form a newly identified protein complex that promotes clearance of Golgi-derived vesicles via the autophagy-lysosome system both in vitro and in vivo, clearly highlighting the role the "Autophagy/Lysosomal degradation" pathway in Parkinson's disease.

More generally, pathway analyses of GWAS data implicated other biological processes as primary etiological events in the disease with significant overrepresentation of association signals in pathways related to "regulation of leucocyte/lymphocyte activity," "cytokine-mediated signaling," "axonal guidance," "focal adhesion," and "calcium signaling."

Representation of genes within each group in the graph is approximate and does not reflect differences in frequency or risk.

Pleomorphic Risk—Exemplified by *SNCA*

This panel illustrates that, at the same locus, several disease-related genetic mechanisms may co-exist, each influencing disease through different biological effects on a single gene. In this particular model, expression of a gene is positively correlated with risk shown by duplication mutations causing Parkinson's disease. Five coding mutations have been identified as the cause of disease in early-onset familial cases. Duplications and triplications of the *SNCA* locus have also been implicated as the cause of early-onset Parkinson's disease. Interestingly, GWAS have also identified two different association signals in this locus, representing common variability with a low effect in the disease. Possible protective variants are also represented in the graph.

ACKNOWLEDGMENTS

Research studies in the authors' lab are mainly supported by Alzheimer's Research UK (ARUK), including a Fellowship to R.G.; by a Fellowship from Alzheimer's Society to J.B.; by the Wellcome Trust/MRC Joint Call in Neurodegeneration award (WT089698) to the UK Parkinson's Disease Consortium, whose members are from the UCL/Institute of Neurology, the University of Sheffield, and the MRC Protein Phosphorylation Unit at the University of Dundee; and by an anonymous donor.

REFERENCES

- Beilina, A., Rudenko, I.N., Kaganovich, A., Civiero, L., Chau, H., Kalia, S.K., Kalia, L.V., Lobbastael, E., Chia, R., Ndukwe, K., et al.; International Parkinson's Disease Genomics Consortium; North American Brain Expression Consortium (2014). *Proc. Natl. Acad. Sci. USA* **111**, 2626–2631.
- Bras, J., Guerreiro, R., and Hardy, J. (2012). *Nat. Rev. Neurosci.* **13**, 453–464.
- Edwards, Y.J., Beecham, G.W., Scott, W.K., Khuri, S., Bademci, G., Tekin, D., Martin, E.R., Jiang, Z., Mash, D.C., French-Mullen, J., et al. (2011). *PLoS ONE* **6**, e16917.
- Holmans, P., Moskvina, V., Jones, L., Sharma, M., Vedernikov, A., Buchel, F., Saad, M., Bras, J.M., Bettella, F., Nicolaou, N., et al.; International Parkinson's Disease Genomics Consortium (2013). *Hum. Mol. Genet.* **22**, 1039–1049.
- Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorf, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., et al. (2009). *Nature* **461**, 747–753.
- Manzoni, C., and Lewis, P.A. (2013). *FASEB J.* **27**, 3424–3429.
- Nalls, M., Pankratz, N., Lill, C.M., Do, C.B., Hernandez, D.G., Saad, M., DeStefano, A.L., Kara, E., Bras, J., Sharma, M., et al. (2014). *Nat. Gen.* **46**, 989–993.
- Singleton, A., and Hardy, J. (2011). *Hum. Mol. Genet.* **20** (R2), R158–R162.